

---

MULTIPLEXED SINGLE-MOLECULE  
FORCE SPECTROSCOPY AND  
ACTIVITY STUDIES ON CELLULOSOMES

---

TOBIAS VERDORFER



München 2018



---

MULTIPLEXED SINGLE-MOLECULE  
FORCE SPECTROSCOPY AND  
ACTIVITY STUDIES ON CELLULOSOMES

---

DISSERTATION

*an der*

LUDWIG-MAXIMILIANS UNIVERSITÄT MÜNCHEN  
FAKULTÄT FÜR PHYSIK  
LEHRSTUHL FÜR ANGEWANDTE PHYSIK

VORGELEGT VON

TOBIAS VERDORFER

*aus Meran*



München, 18.05.2018

Erstgutachter: Prof. Dr. Hermann E. Gaub  
Zweitgutachter: Prof. Dr. Ralf Jungmann  
Tag der mündlichen Prüfung: 12.07.2018





---

## ZUSAMMENFASSUNG

Cellulosome sind große, extrazelluläre Polyprotein-Maschinen, welche lignocellulosehaltige Biomasse auf effiziente Art abbauen. Sie organisieren verschiedene cellulolytische Enzyme auf multivalenten Proteingerüsten (Scaffoldins) mit Hilfe der hochspezifischen Protein-Protein-Wechselwirkungen zwischen Cohesinen und Dockerinen. Cellulosomale Scaffoldins vermitteln zudem Adhäsion zwischen ihrer Wirtszelle und deren Substrat und können bemerkenswert hohen Kräften standhalten.

Aufgrund der synergistischen Organisation von Enzymen sind Cellulosome vielsprechend für industrielle Anwendungen beim Abbau von lignocellulosehaltiger Biomasse zu einfachen Zuckern. Die enorme Vielfalt der hochaffinen Cohesin-Dockerin-Paare ist darüber hinaus von großem Interesse in rational gestalteten künstlichen Enzymnetzwerken, wie Designer-Cellulosomen.

Das Hauptziel dieser Dissertation ist es, die Mechanobiologie hinter der beeindruckenden mechanischen Stabilität cellulosomaler Cohesin-Scaffoldins aufzuklären. Um dies zu erreichen, wurde eine schnelle und parallelisierte Probenpräparationsmethode für multiplex Rasterkraftmikroskop-basierte Einzelmolekül-Kraftspektroskopie (im folgenden AFM-SMFS) entwickelt. Diese Methode ermöglichte die mechanische Charakterisierung aller Cohesindomänen eines enzymtragenden Scaffoldins mit hohem experimentellen Durchsatz und ausgezeichneter Vergleichbarkeit relativer Kraft. Damit konnte die mechanische Stabilität der getesteten Cohesine eindeutig mit ihrer Position innerhalb des cellulosomalen Scaffoldins korreliert werden. Darüber hinaus konnte gezeigt werden, dass Cohesine ihre mechanische Stabilität unter Bindung eines Dockerins deutlich erhöhen. Dies war unerwartet, da sich die Dockerin-Bindungsstelle der Cohesine an der gegenüberliegenden Seite des für die mechanische Stabilität verantwortlichen Hauptelements befindet, dem sogenannten *mechanical clamp motif*. Eine Kombination aus *all-atom* Moleküldynamiksimulationen und multiplex AFM-SMFS ermöglichte die Identifizierung mehrerer stabilisierender Punktmutationen, welche die mechanische Stabilität der schwächsten Cohesin-Domäne mehr als verdoppelten.

Der zweite Teil dieser Dissertation beschäftigt sich mit der Entwicklung eines Assays zur Untersuchung des Cellulose-Abbaus durch Enzymkombinationen und Designer-Cellulosome. Der Assay beruht auf enzymvermittelter Radikalkettenpolymerisation, wodurch der Substratabbau in einem schnellen und parallelen 96-Well-Format zeitaufgelöst detektiert werden kann. Der Assay konnte auf verschiedensten löslichen und unlöslichen Cellulosesubstraten mit hoher Empfindlichkeit durchgeführt werden kann. Darüber hinaus ermöglicht es in Kombination mit Totalreflexions-Fluoreszenzmikroskopie oder AFM Bildgebung eine Echtzeit-Lokalisierung von enzymatischer Aktivität auf Lignocellulose; ein einzigartiges Merkmal unter Cellulase-Aktivitätsassays.

Die Ergebnisse in dieser Dissertation erweitern das Verständnis der Mechanobiologie cellulosomaler Cohesindomänen und des Abbaus von Biomasse durch Enzyme bei. Dies sind zwei zentrale Aspekte für das rationale Design von robusten und effizienten Enzymnetzwerken und Designer-Cellulosomen.



# Contents

ZUSAMMENFASSUNG	vii
CONTENTS	ix
<b>I Introduction</b>	<b>1</b>
<b>II Scientific Context</b>	<b>7</b>
1 THE CELLULOSOME - A MODEL PROTEIN NANOMACHINERY	9
1.1 Lignocellulosic Biomass . . . . .	9
1.2 Free Cellulolytic Enzymes and Cellulosomes . . . . .	10
1.2.1 <i>Trichoderma reesei</i> : A Model Organism for Free Cellulases . .	12
1.2.2 <i>Acetivibrio cellulolyticus</i> : A Model Organism for Cellulosomes	13
1.3 The Remarkable Mechanical Stability of the Cohesin-Dockerin Interaction . . . . .	15
1.4 Designer Cellulosomes . . . . .	15
1.5 Assaying Lignocellulose Degradation . . . . .	16
2 AFM-BASED SMFS ON COHESINS AND DOCKERINS	19
2.1 The Basics of AFM-based SMFS . . . . .	19
2.1.1 Sample Preparation for AFM-based SMFS . . . . .	21
2.1.2 The Worm-Like Chain Polymer Elasticity Model . . . . .	22
2.1.3 The Bell-Evans Model for Protein Unfolding and Receptor-Ligand Unbinding . . . . .	23
2.2 AFM Cantilever Calibration . . . . .	23

<b>III Results</b>	<b>27</b>
3 PEER-REVIEWED RESEARCH ARTICLES	29
3.1 PUBLICATION P1: Combining <i>in Vitro</i> and <i>in Silico</i> Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics . . . . .	29
3.2 PUBLICATION P2: Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy . .	69
3.3 PUBLICATION P3: From Genes to Protein Mechanics on a Chip . . . .	85
3.4 PUBLICATION P4: Biasing Effects of Receptor-Ligand Complexes on Protein-Unfolding Statistics . . . . .	107
3.5 PUBLICATION P5: Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity . . . . .	117
3.6 PUBLICATION P6: Quantifying Synergy, Thermostability, and Targeting of Cellulolytic Enzymes and Cellulosomes with Polymerization-Based Amplification . . . . .	131
<b>IV Conclusion</b>	<b>143</b>
<b>V Appendix</b>	<b>147</b>
LIST OF PUBLICATIONS	149
LIST OF FIGURES	151
BIBLIOGRAPHY	153
ACKNOWLEDGEMENTS	165





**Part I**

**Introduction**



---

Proteins are fundamental to living organisms and therefore a core subject of biological and biophysical research. They perform countless vital functions within and outside their hosts, including replication, metabolic reactions, signal transduction, molecule transport or structural and mechanical tasks [1–5]. In the majority of cases, elaborate macromolecular assemblies or protein networks synergistically perform complex tasks or reaction cascades. Investigating individual proteins and their activity at a single-molecule level is often imperative to understand their intricate underlying mechanisms of action.

The investigations in this thesis were inspired by large, complex multi-protein systems called cellulosomes, which synergistically coordinate enzymes with different specificities and non-catalytic domains [6]. Cellulosomes are secreted by various anaerobic bacteria and fungi for the efficient degradation of lignocellulosic biomass. The key players in cellulosome assembly are multimodular adapter proteins called scaffoldins, which mainly consist of cohesin domains. Enzymes and non-catalytic auxiliary modules carry dockerin domains, which can bind to cohesins. Through this high-affinity receptor-ligand interaction between cohesins and dockerins, scaffoldins enable a spatially synergistic arrangement of enzymes and non-catalytic domains, as well as robust cell adhesion to cellulose in mechanically stressful environmental conditions [7]. Due to these intriguing features, cellulosomal scaffoldins are particularly interesting as building blocks in rationally designed artificial enzyme networks, such as designer cellulosomes.

THE MAIN AIM OF THIS THESIS IS TO characterize the mechanical properties of an entire cellulosomal scaffoldin, which has never been achieved before. Atomic force microscopy-based single-molecule force spectroscopy (AFM-based SMFS) is the ideal tool to understand the mechanobiology that governs the remarkable robustness of cellulosomal components. It was used previously to characterize the non-covalent protein-protein interaction between cohesins and dockerins, as well as the mechanostability of cohesin domains themselves [8–11]. Classical AFM-based SMFS, however, typically lacks the required throughput or absolute force precision to achieve complete and comparative mechanical phenotyping of cohesin scaffoldins, or more general: protein libraries.

**Publication P1** and **publication P2** tackle these limitations by employing a newly developed multiplexed, cell-free *in vitro* protein expression and site-specific pulldown sample preparation protocol for AFM-based SMFS. This method enabled the mechanical characterization of all cohesin domains from an enzyme-bearing scaffoldin with unprecedented experimental throughput and excellent relative force comparability. I was thus able to clearly correlate the mechanical stability of the tested cohesins with their positions within the cellulosomal scaffoldin. Furthermore, I found that all investigated cohesins notably increase their mechanical stability upon binding of a dockerin, although the cohesin's binding interface is located at the opposite side of its main element responsible for mechanical stability, the mechanical clamp motif. A combination of all-atom steered molecular dynamics simulations and multiplexed AFM-based SMFS further contributed to the understanding of cohesin mechanostability and enabled the identification of several mechanically stabilizing point mutations, one of which increased the mechanical stability of the weakest cohesin domain by more than a factor of two.

---

**Publication P3** further incorporates the aforementioned cell-free protein expression and immobilization protocol into an versatile microfluidic platform. This lab-on-a-chip approach can in principle be used to prepare hundreds of AFM-SMFS protein samples in parallel, allowing for high-throughput mechanical phenotyping of large protein libraries.

**Publication P4** presents a model-free theoretical framework for biasing effects which can occur in the single-molecule force spectroscopy unfolding measurements of high-strength protein domains. It predicts and corrects biasing of overlapping force distributions and provides a framework for the right choice of pulling handle in SMFS experiments.

THE SECOND PART OF THIS THESIS is the development of a versatile assay for the investigation of lignocellulose degradation by enzyme compositions and designer cellulosomes. Typical cellulase activity assays have strict limitations when used in high-throughput applications and do not allow for enzyme localization on natural substrates.

**Publication P5** and **publication P6** present a novel cellulase activity and localization assay which relies on enzyme-mediated radical chain polymerization for strong signal amplification. It allows optical, time-resolved detection of substrate degradation with high sensitivity in a fast and parallel 96-well format on a wide range of different soluble and insoluble cellulose substrates. More importantly, it allows real-time localization of enzymatic activity when combined with total internal reflection fluorescence microscopy or AFM imaging, a unique feature among cellulase activity assays.





**Part II**

**Scientific Context**



# 1

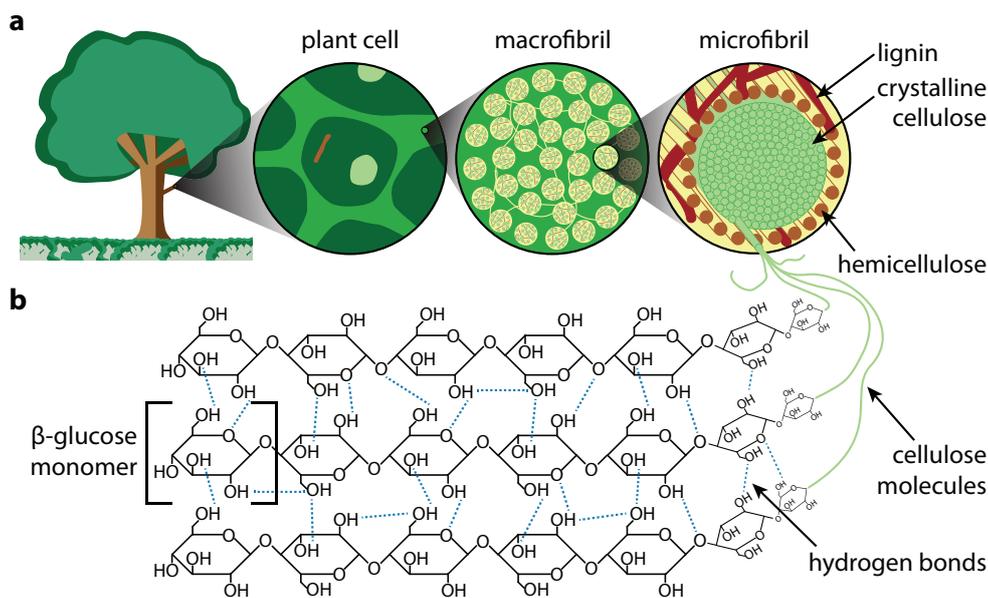
## The Cellulosome - a Model Protein Nanomachinery

The following sections give an overview over the recalcitrant structure of lignocellulosic biomass, the strategies used by cellulosomes to efficiently hydrolyze it and the central role of scaffoldins within cellulosomes. A short description of the function of different cellulolytic enzymes, non-catalytic domains and cellulosomal components is given at the example of two cellulose-degrading model organisms: the fungus *Trichoderma reesei* and the bacterium *Acetivibrio cellulolyticus*; each representing one of the two paradigms of cellulase organization. Furthermore, the concept of designer cellulosomes is shortly introduced and recent advances therein are summarized.

### 1.1 Lignocellulosic Biomass

Cellulose is the most abundant renewable biopolymer and major reservoir of fixed carbon on earth. Besides hemicellulose and lignin, it is the main constituent of plant cell walls, but is also found in algae and bacteria [12]. Long linear chains of several hundred to thousands of glucose molecules give cellulose its crystallinity, strength and resistance to hydrolysis, making it the main structural component of the primary plant cell wall. Hemicellulose, in contrast, is chemically not well defined but rather a family of polysaccharides. It has a random, amorphous and branched structure with little strength and is easily hydrolyzed by dilute acid and base. Its main job is to cross-link cellulose nanocrystals in the plant cell wall through hydrogen bonds and hydrophobic interactions. Lignin is a class of complex organic polymers which provide flexibility and resistance to compression by embedding cellulose and hemicellulose in an amorphous matrix (i.e., lignocellulose). Lignin consists of cross-linked phenolic polymers making it highly resistant to acid- and base-catalyzed hydrolysis [13].

Typical proportions in plant biomass are 30-50% cellulose, 20-35% hemicellulose and 10-30% lignin, but strongly depend on its source [12, 14]. Due to the high proportion of cellulose and hemicellulose, lignocellulose is generally considered to be the most promising renewable source for a sustainable production of biofuel *via* fermentation of glucose, with major existing reservoirs in the form of agricultural waste [15].



**Figure 1.1. Structure of lignocellulose.** (a) Plant cells consist of a hierarchical organization of cellulose. Crystalline cellulose is cross-linked by hemicellulose and lignin into structures called microfibrils, which are further organized into macrofibrils. This complex composition results in high structural integrity and flexibility in plant cell walls. (b) Cellulose is the main component of lignocellulose and consists of glucose molecules linked by 1,4- $\beta$ -glycosidic bonds. Hydrogen bonds between the different layers of polysaccharides give cellulose its crystallinity and resistance to degradation.

The basis for this figure was kindly provided by Wolfgang Ott.

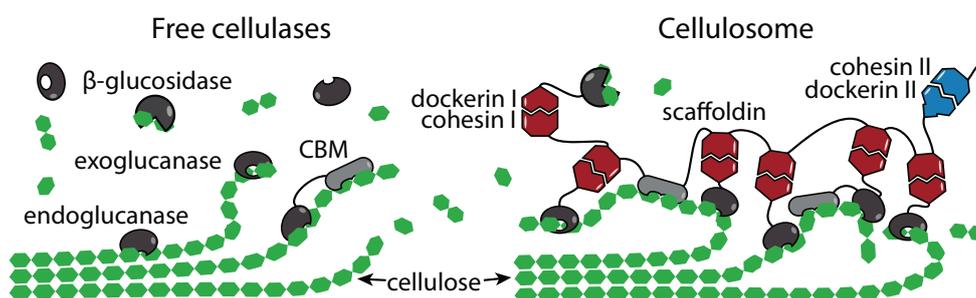
The combination of crystalline and heterogeneous compounds, and the high degree of lignification gives lignocellulose its dense structure and chemical stability. The industrial degradation of lignocellulosic polysaccharides into their component carbon chain sugars therefore heavily depends on acid or heat treatments [16], making this process expensive, slow, relatively inefficient and environmentally damaging [17]. A variety of organisms, however, have evolved to efficiently degrade lignocellulose and take advantage of this nutrient source. Extensive studies on a handful of model organisms such as the fungus *Trichoderma reesei* [18], and the bacteria *Clostridium thermocellum* [19] and *Acetivibrio cellulolyticus* [20] have shown that the strategies for enzymatic biomass degradation are diverse and complex. Numerous cellulolytic enzymes with different activities in combination with non-catalytic adapter and auxiliary domains are employed to synergistically and efficiently degrade the chemically and structurally recalcitrant substrate.

## 1.2 Free Cellulolytic Enzymes and Cellulosomes

The two paradigms of enzymatic lignocellulose degradation are free enzyme systems and cellulosomes (Figure 1.2). In both cases different cellulolytic enzymes (i.e. cellulases) are employed to extract soluble carbohydrates from plant cell walls. Most importantly, cellulases include exo- and endo-acting enzymes and  $\beta$ -

glucosidases, which work synergistically to hydrolyze the recalcitrant lignocellulose substrate and crystalline cellulose microfibrils.

Exoglucanases processively degrade cellulose from its chain ends, while endoglucanases randomly hydrolyze glucosidic bonds within the polysaccharide chain [21]. Finally,  $\beta$ -glucosidases cleave cellobiose, a dimer of  $\beta$ -glucose molecules, into glucose and thus counteract exo- and endo-cellulase product inhibition. To be able to target specific substrates, cellulases are often accompanied by non-catalytic protein domains. Carbohydrate binding modules (CBMs), for example, target specific forms of cellulose and thus guide catalytic protein domains to their substrate [22].



**Figure 1.2. The two paradigms of cellulase organization.** Free enzymes systems let the cellulolytic enzymes diffuse freely, which can lead to relatively inefficient cellulose hydrolysis. Cellulosomes typically coordinate cellulolytic enzymes and auxiliary domains, like carbohydrate binding modules (CBMs), on large multi-domain protein scaffoldins through high-affinity cohesin-dockerin interaction with different binding specificities (red and blue). Additionally, scaffoldins anchor the host cell to its substrate. This results in synergistic cellulose degradation and constant substrate proximity.

In contrast to the free enzyme systems, which lack the ability to synergistically organize their constituents and to retain proximity to the substrate, cellulosomes utilize large non-catalytic adapter proteins called scaffoldins to organize and localize auxiliary domains and cellulases with various types of action, and to assure proximity between the host cell and its substrate. The assembly of the frequently complex, branching cellulosome architectures is dictated by the specific, high-affinity interaction between cohesins and dockerins of different types or different affinity profiles [23, 24]. Their interaction is usually species-specific, thus, a cohesin of one species will typically only recognize and bind the dockerins from the same species, but not the dockerins from a different species [25]. Furthermore, cohesin-dockerin pairings are typically classified into three types, according to their position and function within the cellulosome. The cohesin-dockerin type-I interaction typically binds enzymes to the cellulosomes enzyme-bearing scaffoldins (i.e., primary scaffoldins), while type-II and type-III typically connect primary to adapter scaffoldins (i.e., secondary, or sometimes even tertiary scaffoldins) or anchor the cellulosome to its host. While all types show high affinities in the picomolar to low nanomolar range, their binding forces differ strongly (see section 1.3) [8, 9, 26, 27].

In the following sections, two popular model organisms, which were used within the research of this thesis, were chosen to represent each of the cellulase organization paradigms and are shortly introduced. The fungus *T. reesei* is an industrially used cellulase hyperproducer and represents the free enzyme system. A commercially available enzyme mix harvested from *T. reesei* was used to characterize the efficiency of the cellulase activity assays in publication P5 and publication P6. The cellulosome system is represented by *A. cellulolyticus*, a bacterium capable of producing and maintaining exceptionally complex and diverse cellulosome arrangements in mechanically demanding environmental conditions. Its intricate strategy of enzyme organization is of high interest in designer cellulosome research and the structural robustness of its primary scaffoldin ScaA was the scope of research in publication P1 and publication P2.

### 1.2.1 *Trichoderma reesei*: A Model Organism for Free Cellulases

The mesophilic fungus *T. reesei* was first isolated from rotting US Army equipment during World War II [28] and its outstanding ability to degrade native crystalline cellulose and lignocellulose was soon utilized in industrial applications [29]. Since its secreted enzymes are not cell-bound, it is possible industrially to harvest large amounts from *T. reesei*'s growth medium supernatant. Its enzymes are used in food, detergent, textile, pulp and paper and biorefinery applications, especially in biomass conversion [30–32]. The cellulase composition produced by *T. reesei* is today's industry standard against which all the improved cocktails are compared [30]. Full genome sequencing and analysis in 2008 revealed that *T. reesei* carries genes for 200 cellulases and 36 CBMs [33].

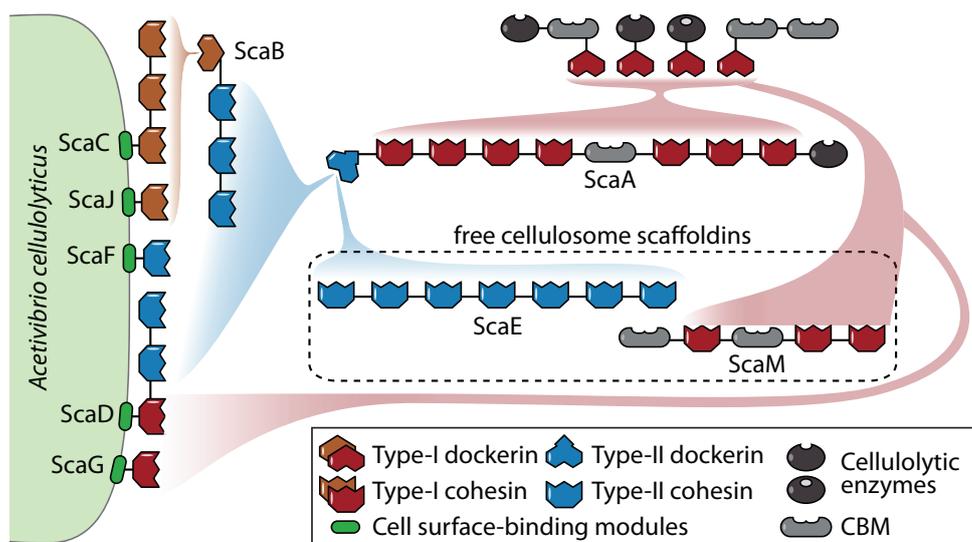
Its most prominent exocellulase CBH1 is of specific importance, since it was the first eukaryotic cellulase to be cloned and the first cellulase whose structure was solved [34, 35]. Furthermore, CBH1 was imaged during cellulose degradation using high-speed AFM in 2011 by Igarashi *et al.* and it was shown that it slides unidirectionally along the cellulose surface but at one point exhibited collective halting analogous to a traffic jam. Treatment of the crystalline cellulose with *T. reesei* CBH2 resulted in a remarkable increase in the proportion of mobile CBH1 molecules on the surface and the cellulose was completely degraded by the synergistic action between the two enzymes [36].

Through mutagenesis programs *T. reesei* was made a cellulase hyperproducer with extracellular protein production reaching up to 100 g/L [37, 38] and was mutated to produce additional enzymes, like  $\beta$ -glucosidases, to achieve more efficient and complete degradation of lignocellulose into fermentable sugars [39–48]. Cellulases derived from *T. reesei* are nowadays involved in roughly 80% of the cellulosic biofuel production world-wide and are readily available for purchase for use in scientific studies. The research on *T. reesei* has laid the groundwork for the current understanding of the synergistic combination of different cellulases and their regulation [28, 49–51] and *T. reesei* remains of high interest in research with more than 100 research articles dealing with the fungus or its enzymes published each year [52].

1.2.2 *Acetivibrio cellulolyticus*: A Model Organism for Cellulosomes

*A. cellulolyticus* is a gram-negative, anaerobic bacterium, which is found in mechanically challenging environments such as municipal sewage sludge [53] or the bovine rumen [54]. Its ability to efficiently degrade different kinds of crystalline cellulose and lignocellulose was first found by Patel *et al.* in 1980 [55] and further studied by MacKenzie *et al.* in 1985 [56].

It was later found by selective, large-scale screening of cohesin-dockerin interactions that *A. cellulolyticus* expresses an exceptionally complex cellulosome system and possesses an intriguing diversity of modular cellulosomal components [20] (see figure 1.3). Containing a total of 41 cohesins and 143 dockerins, *A. cellulolyticus* has one of the largest number of potential cohesin-dockerin interactions of any organism. Some unusual and novel features of *A. cellulolyticus*' cellulosome are a number of scaffoldins with unusual cohesin-dockerin arrangements and cellulolytic components fused to normally non-cellulosomal proteins such as peptidases, dockerin dyads or domains with unknown function [20, 57–59]. The currently known cellulosome architecture of *A. cellulolyticus* consists of 16 scaffoldin proteins and three orthogonal cohesin-dockerin binding profiles, but only about 7% of all its theoretical cohesin-dockerin interactions have to date been screened [20].



**Figure 1.3. Simplified architectural model of *A. cellulolyticus*' cellulosome system.**

Even with only the most prominent and important scaffoldins shown, *A. cellulolyticus*' cellulosome system is still surprisingly complex. The color coding of the cohesins and dockerins (brown, blue and red) reflect their different, orthogonal affinity profiles. Several scaffoldins (ScaC, ScaJ, ScaF, ScaD and ScaG) are anchored in *A. cellulolyticus*' cell wall and build the basis for its cell-bound cellulosome system. Scaffoldins ScaF and ScaD can directly anchor *A. cellulolyticus*' main enzyme-binding scaffoldin ScaA to its cell wall, while ScaC and ScaJ only bind the adapter scaffoldin ScaB. Scaffoldin ScaA contains a cellulose-binding module (CBM) and a catalytic domain, which is atypical for cellulosomal scaffoldins. Scaffoldins ScaE and ScaM are not bound to their host cell and represent the basis for *A. cellulolyticus*' free cellulosome system. A full representation of *A. cellulolyticus*' cellulosome system can be found in Hamberg *et al.* (2014) [20].

Interestingly, *A. cellulolyticus*' cellulosome consists of both several different cell-bound architectures, as well as of soluble cellulosome systems that are not bound to the bacterial cell surface. Both systems are well capable of assembling large cellulosomal network. If, for example, three ScaB scaffoldins bind to a single ScaC, and three ScaAs bind to each of the ScaBs, a total of 72 cellulolytic enzymes will be orchestrated onto *A. cellulolyticus*' cell wall at a single anchor. Seven ScaAs will assemble 56 enzymes onto one free-floating ScaE.

In both systems though, scaffoldin ScaA is almost exclusively responsible for the attachment of cellulolytic enzymes to the cellulosome. At the same time, ScaA has the essential function of targeting specific substrates and anchoring the cell to the cellulose fibril *via* its CBM. Scaffoldin ScaA therefore serves a dual purpose for the organism by incorporating catalytic modules into the cellulosome complex and by adhering the bacterium to the cellulosic substrate, highlighting the central role of this scaffoldin for *A. cellulolyticus*' cellulosome. ScaA's ability to remain folded and thus functional under mechanical stress thus strongly affects *A. cellulolyticus*' cellulosome functionality.

Although all cohesin domains from scaffoldin ScaA show the same binding specificity to dockerins and a very high sequence similarity among each other, they fulfill different mechanical tasks and can thus be classified into two distinct groups. The cohesins located between ScaA's type-II dockerin, which connects the scaffoldin to the cell, and its CBM form the so-called "bridging" region of the scaffoldin. The cohesins located exterior from the CBM, on the other hand, form the "hanging" region of this scaffoldin. The cohesin domains located within the bridging region are expected to be mechanically stressed in turbulent environments, since the CBM anchors the cell to its lignocellulose substrate. It was therefore hypothesized that the cohesins of a scaffoldins bridging region should be able to withstand higher mechanical loads, in order to keep the cellulosome and its assembly intact and functional. This hypothesis has previously been successfully for a limited selection of cohesins.

Valbuena *et al.* investigated the mechanical stability of two cohesins from *C. thermocellum*'s scaffoldin CipA, one taken from the scaffoldins bridging region and one taken from its hanging region [11]. While the cohesin module from the hanging region unfolded at around 215 pN, the cohesin module from the bridging region of scaffoldin CipA was able to withstand 480 pN, which placed it among the most robust mechanical proteins studied experimentally at the time. Valbuena *et al.* attributed the differing mechanical strength of the cohesins to the cohesins ability to form a so-called mechanical clamp motif, a patch of backbone hydrogen bonds between parallel N- and C-terminal  $\beta$ -sheets.

Since mechanical stability is an important prerequisite for viable candidate cohesins in designer cellulosomes and other rationally designed, cohesin-dockerin-based enzyme networks, all cohesins from *A. cellulolyticus*' primary scaffoldin ScaA were the subject of investigation in publication P1 and publication P2.

### 1.3 The Remarkable Mechanical Stability of the Cohesin-Dockerin Interaction

The protein-protein interaction between cohesins and dockerins is a fascinating system with diverse applications, since it combines high specificity, affinity and mechanical strength. Due to these important characteristics it has been the subject of several AFM-based SMFS studies over the last years.

Stahl *et al.* reported on the first mechanical study on cohesin-dockerin unbinding in 2012, using cohesin 2 from CipA and the dockerin from Cel48S from *C. thermocellum* [8]. Utilizing an experimental design with two so-called molecular fingerprints, it was possible to unambiguously identify single, specific cohesin-dockerin unbinding events. The measured rupture forces of 100 - 150 pN at loading rates of 0.8 - 20 nN/s were, at the time, among the highest receptor-ligand strengths reported. They observed force-induced dissociation of calcium ions from the dockerins loop-helix F-hand motifs, which in the presence of EDTA resulted in loss of affinity to the cohesin partner. Furthermore, it was observed that the cohesin-dockerin interaction unbinds in either one or two distinct final force steps, which in combination with the high structural symmetry of the dockerin hinted towards the existence of a dual binding mode between the two binding partners. The same cohesin-dockerin pair was used as specific pulling handle in publication P1.

It was later shown by Jobst *et al.* that the same type-I cohesin-dockerin interaction exhibits two distinct modes of binding [10] with equal probabilities, but notably differing unbinding forces. This finding was made possible by exploiting a usually unwanted effect in single-molecule force spectroscopy measurements, called fingerprint biasing. It is caused by overlapping force probability distributions of molecular fingerprints and pulling handles, and enabled a clear distinction between two receptor-ligand binding modes and two unfolding pathways. A theoretical framework for this effect is presented in publication P4.

Schoeler *et al.* found in 2014 that the type-III ScaE-CttA cohesin-dockerin interaction from *Ruminococcus flavefaciens*, which is solely responsible for maintaining bacterial adhesion through two CBM domains, is able to withstand impressive unbinding forces of 600 - 750 pN at loading rates of 10 - 100 nN/s [9]. It was later shown through all-atom steered molecular dynamics simulations and network-based correlation analysis that this incredible mechanical stability can be attributed to inter-domain complex stabilization and force propagation pathways nonparallel to the pulling axis [60]. Since then *R. flavefaciens*' type-III cohesin-dockerin interaction has become a popular pulling handle in AFM-based SMFS experiments, due to its high specificity, high unbinding forces and long-term stability [61–63], and was also used in publication P1 and publication P2.

### 1.4 Designer Cellulosomes

The concept of designer cellulosomes was first introduced by Bayer *et al.* in 1994 [64] and a first proof-of-concept was accomplished by Fierobe *et al.* in 2002 [65]. Designer cellulosomes utilize cellulosomal components from different organisms and combine them to further increase cellulolytic efficiency.

The construction of designer cellulosomes can be achieved in several different

ways. One approach is to incorporate one or several new enzymes into an existing cellulosome system, in order to alter its substrate specificity or improve its performance. Since the overall structure of the cellulosome remains unchanged, all of its benefits like substrate targeting and synergistic enzyme coordination remain intact [66]. Gefen *et al.*, for example, designed a chimeric cohesin-fused  $\beta$ -glucosidase that binds directly to the *C. thermocellum*'s cellulosome through an unoccupied dockerin module of its major scaffoldin subunit [67]. This resulted in higher overall cellulose degradation compared to the native cellulosome alone or in combination with wild-type  $\beta$ -glucosidase in solution.

A more sophisticated and ambitious idea is to substitute all native cellulosomal components, apart from the scaffoldins, and replace them with enzymes and auxiliary domains more suitable for the desired task. This concept has been successfully tested by Cho *et al.* in 2004 [68]. The remaining downside to this approach is that the position and distribution of cellulases on the primary scaffoldin cannot be controlled.

The only way to improve this concept further is by designing so-called chimeric scaffoldins from scratch, by combing different cohesin domains with orthogonal binding specificities, and desired thermal or mechanical properties. This enables the precise incorporation of complementary dockerin-containing components into the complex. This concept to date been successfully tested for chimeric scaffoldins with two to eight orthogonal cohesin domains [21, 69–73].

The concept of designer cellulosomes on chimeric scaffoldins can of course easily be adapted for other enzymatic systems which profit from proximity effects. Basically any imaginable combination of enzymes and non-catalytic domains can be recombinantly combined with dockerin domains and placed onto cellulosomal cohesin scaffoldins through the high-affinity interaction between cohesins and dockerins. The individual components of such rationally designed protein machineries can either be produced and secreted by a single organism, or individually expressed in suitable surrogate organisms, purified and combined *in vitro*. This concept has been shown to increase the production rate of several different multi-enzyme reaction cascades by up to an order of magnitude, due to efficient substrate channeling and increased reaction rates [74–77].

## 1.5 Assaying Lignocellulose Degradation

In order to improve the hydrolysis efficiency of enzyme formulations and designer cellulosomes, suitable tools for the quantification and general investigation of cellulose degradation are required.

Popular cellulase activity assays, such as the IUPAC-standardized filter paper assay (FPA) or the glucose oxidase/horseradish peroxidase (GOx/HRP) system, have distinct drawbacks when used in high-throughput screening [78]. Often can they not be applied to detect cellulolytic efficiency on insoluble, real-life lignocellulose substrates, and typically only quantify the total amount glucose released from lignocellulose degradation [79]. Another large drawback of the widely used FPA and some other activity assays is that they can not be used to detect the release of glucose in real time, but only as endpoint measurements, due their strong acidic, alkaline, thermal or other physio-chemical reaction conditions,

which are incompatible with with cellulolytic enzymes [80].

The spatial localization of cellulase hydrolysis activity on topologically complex lignocellulose can provide new insights into enzymatic activity and synergy, but has thus far been difficult to achieve. Conventional high-resolution imaging methods such as transmission electron microscopy, scanning electron microscopy or AFM imaging have significant drawbacks, which deny the monitoring of enzymatic activity under biocompatible conditions or on real lignocellulosic substrates [79, 81, 82].

The method presented in publication P5 and publication P6 addresses some of the shortcomings of the aforementioned cellulase activity and localization assays. It utilizes enzyme mediated radical chain polymerization, which forms a localized fluorescent hydrogel in response to released glucose, and can be used in a 96-well plate format to quantify cellulase activity on a wide range of soluble or insoluble substrates. Furthermore, it can be combined with high resolution imaging methods - either total internal reflection fluorescence microscopy or AFM imaging - for the localization of enzymatic activity on lignocellulose in real time. A patent application for this assay technology was granted in 2015 (international patent number WO2015091772 A1) [83].



# 2

## AFM-based SMFS on Cohesins and Dockerins

Single-molecule force spectroscopy allows to probe the unfolding of single protein domains, the unbinding of receptor-ligand interactions or the elastic response of polymer strings like DNA [84–88]. The AFM is uniquely capable of probing forces from several pN up to  $\sim\mu\text{N}$ , while still achieving sub-nanometer spatial resolution [89], making it ideal for the mechanical characterization of cellulosomal components.

This chapter gives an overview over the principles of AFM-based SMFS and the most important theoretical models for data analysis used in the investigations within publication P1, publication P2 and publication P3.<sup>1</sup>

### 2.1 The Basics of AFM-based SMFS

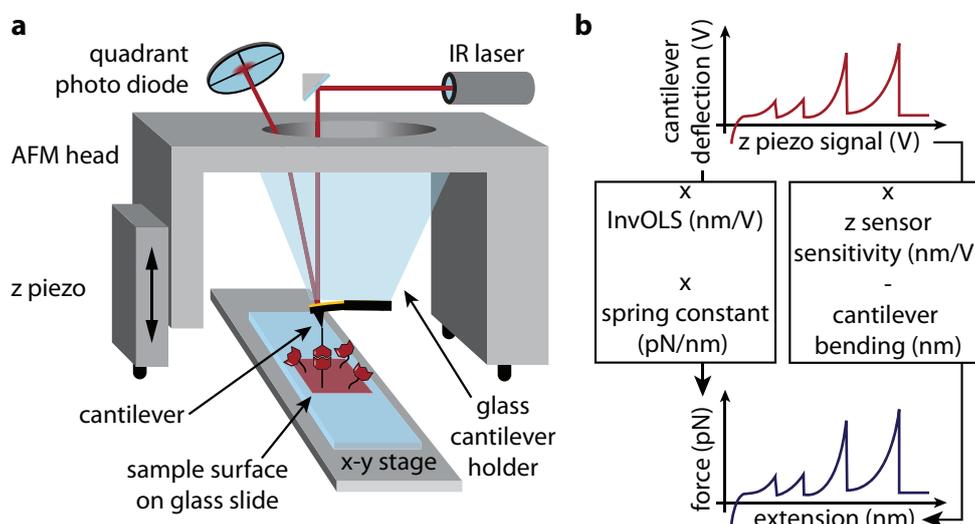
The atomic force microscope relies on a spring-like cantilever with a sharp tip at its end. While the size of the cantilever itself is on the order of micrometers, the tip radius is typically on the order of nanometers. The AFM was originally invented as a type of scanning probe microscope by Binnig, Quate and Gerber in 1986, which can be used on non-conductive samples, and achieved resolutions on the order of fractions of nanometers, thus well below the optical diffraction limit [92]. Since the vertical bending of a cantilever can easily be detected and converted into force units by multiplication with its spring constant, the AFM was quickly applied to investigate the elastic and viscoelastic properties of samples by nanoindentation [93–95] and to quantify mechanical forces acting between the tip of the cantilever and a sample, such as receptor-ligand interactions or chemical bonds [84–86, 96–98].

A typical AFM-based SMFS setup is shown in figure 2.1 (a). The head of the AFM is placed over a sample, which can be moved *via* an x-y stage. The vertical bending of the cantilever under force (i.e., cantilever deflection) is most commonly determined *via* the reflection of a focused infra-red laser beam off the cantilevers reflective backside onto a quadrant photodiode, resulting in typical force noise

---

<sup>1</sup>A complete description of the experimental and theoretical procedures in AFM-based SMFS can be found in the doctoral thesis of Markus Jobst [90]. A full overview over the current capabilities and possibilities of AFM-based SMFS can be found in the review article by Ott *et al.* [91].

from thermal fluctuations on the order of few piconewtons. The AFM head itself, and thus the cantilever relative to the surface, can be moved using a z piezoelectric actuator with sub-nanometer precision.



**Figure 2.1. Schematic of an AFM in single-molecule force spectroscopy mode. (a)** The cantilever is mounted on a glass holder and an infrared laser is reflected off its backside. The bending of the cantilever under force is detected *via* the position of the reflected laser beam on a quadrant photodiode. The AFM head itself is moved with sub-nanometer precision *via* a z piezo. The sample under the AFM is moved in between measurements *via* an x-y stage. **(b)** The raw data signal from a constant speed AFM-SMFS pulling measurement consists of cantilever deflection volts *vs.* z piezo signal volts. A typical data curve from a constant speed SMFS measurement shows a sawtooth-like pattern, where peaks correspond to the unfolding of individual protein domains, and the final peak corresponds to the unbinding of the pulling handle. The cantilever deflection volts are converted into force units by multiplication with the inverse optical lever sensitivity and the cantilever's spring constant. The z piezo signal is converted into length units by multiplying it with the z sensor sensitivity and by subtracting a correction factor for the bending of the cantilever under force. The basis for this figure was kindly provided by Jochen P. Müller and Magnus S. Bauer.

In a typical receptor-ligand-based AFM-SMFS experiment the cantilever is brought into contact with the surface, where immobilized molecules will stick to the tip through a pulling handle, for example a cohesin-dockerin interaction. Upon retraction of the cantilever with constant speed *via* the z piezo, force is applied to the polymer string and the cantilever bends proportionally to it. A typical constant-speed data curve shows a sawtooth-like pattern, where peaks correspond to the unfolding of individual protein domains, and the final peak corresponds to the unbinding of the pulling handle. So-called molecular fingerprint domains with known unfolding behavior are often employed, which allow a clear identification of single interaction.

The actual data acquisition process is schematized in figure 2.1 (b). Initially, cantilever deflection volts *vs.* z piezo volts are recorded as raw data. The cantilever deflection volts are converted into force units by multiplication with

the inverse optical lever sensitivity (InvOLS) and the cantilevers spring constant. The acquisition of these two vital parameters - and therefore the force calibration itself - can be prone to considerable errors, which makes comparative AFM-SMFS studies with measurements from individual cantilevers difficult. The details of cantilever calibration are briefly discussed in section 2.2. The actual distance between the tip of the lever and the surface is calculated by multiplying the z piezo voltages with the piezo's sensitivity, followed by the subtraction of a term correcting for the bending of the cantilever under force. The resulting force *vs.* extension curve is then used to extract the mechanical parameters of the unbinding of a receptor-ligand interaction or of the unfolding of a protein domain.

### 2.1.1 Sample Preparation for AFM-based SMFS

The classical way of preparing a sample for AFM-based SMFS is to bind purified proteins to a single chemically functionalized spot on a glass slide. The proteins are typically obtained by transforming surrogate bacteria (e.g. *Escherichia coli*) with the plasmid DNA coding for the desired protein. After identification of successful transformation, the bacteria are incubated to overexpress the protein constructs, which can be obtained from the cells through lysis and subsequent protein purification. This process - from DNA to purified protein and AFM-SMFS sample preparation - typically takes several days and thus marks a significant bottleneck in the multiplexed mechanical phenotyping of protein domains.

The methods for protein expression and sample preparation in publication P1, publication P2 and publication P3 utilize cell-free *in vitro* translation and transcription (IVTT) reaction kits to express the desired proteins and bind them to a functionalized surface *via* an additionally supplemented enzyme in a one-pot reaction. IVTT reaction kits can either be crude, harvested cell extracts or bottom-up reconstituted protein synthesis kits, where all necessary components needed for *in vitro* transcription and translation are individually purified from *E. coli* and then recombined. The obvious advantage of IVTT kits is the significantly shorter protein expression time of only several hours, and the ability to execute small-scale reaction as small as microliters. Typical disadvantages are their relatively high cost and comparatively low protein expression yields. However, since volumes in the  $\mu\text{l}$  range and protein concentrations in the  $\mu\text{M}$  range are ideal for the preparation of a typical AFM-SMFS sample surface, these drawbacks become obsolete.

While publication P1 and publication P2 use a microwell mask to compartmentalize the surface of a microscope glass slide for parallelized sample preparation, publication P3 utilizes an intricate microfluidic chip to do so. The advantages of the microwell mask over the microfluidic chip are that it can be obtained from standard lab suppliers, that the sample preparation workflow and liquid handling are relatively simple and can be done with standard lab equipment, and that the alignment of the AFM cantilever to the millimeter-sized spots can in principle be done by the bare eye. Its use is thus adequate for the comparative AFM-based SMFS study of up to ten proteins with a single cantilever.

The microfluidic chip, on the other hand, is in principle capable of preparing up to 640 individual protein spots, but the in-house fabrication of the chips themselves, as well as the sample preparation workflow and handling require special equipment and training. The alignment of the cantilever with the relatively

small protein spots (diameter  $\sim 100 \mu\text{m}$ ) furthermore require the use of a combined TIRF-AFM setup.

### 2.1.2 The Worm-Like Chain Polymer Elasticity Model

Since the force extension behavior of linear polymers like proteins in water is predominantly entropy driven, it is essential to use suitable polymer elasticity models to extract information from force-extension data. The worm-like chain (WLC) model was first introduced by Kratky and Porod in 1949 and is most commonly used in the data analysis of protein-based AFM-SMFS measurements [99]. It approximates the backbone of a protein as a continuous, linear polymer chain, where the correlation between two tangent vectors drops exponentially with the distance  $\Delta s$  between them:

$$\langle \vec{t}(s) \cdot \vec{t}(s + \Delta s) \rangle = \exp\left(-\frac{|\Delta s|}{l_p}\right). \quad (2.1)$$

with  $l_p$  being the so-called persistence length of the polymer, where the directional correlation drops to  $\frac{1}{e}$ . Stiffer polymers thus have a larger persistence length than softer ones. The persistence length of double-stranded DNA, for example, was determined to be around 35 nm [100], while the persistence length of a typical protein backbone is around 0.365 nm [101].

Forcing the ends of a polymer chain with contour length  $l_c$  into an end-to-end separation  $x$  reduces its conformational space and thus results in an entropically driven restoring force. Since no analytical solution for the relation between end-to-end distance and the resulting force of the WLC model is known, several approximations were found over the years. The most used approximation was introduced by Marko *et al.* in 1995 [102]:

$$F_{\text{WLC}}(x) = \frac{k_B T}{l_p} \left( \frac{x}{l_c} + \frac{1}{4(1 - \frac{x}{l_c})^2} - \frac{1}{4} \right) \quad (2.2)$$

Over the years more accurate approximations [103] and extended models which include quantum mechanical corrections for the stretching of the protein backbone at high forces were found [104].

The WLC model is a vital tool in AFM-SMFS data acquisition, since it can be used to transform force-distance traces into contour length space. Each of the stretches in a typical sawtooth-like unfolding pattern of several protein domains would thus ideally result in a single peak in contour length space. The distance, or contour length increment, between said peaks corresponds to the hidden stretch of protein backbone within a folded protein domain, which is released when the domain is unfolded. So-called fingerprint domains with known contour length increments and unfolding behavior can be incorporated into the polyprotein constructs and thus allow for the identification of specific, single interactions.

## 2.1.3 The Bell-Evans Model for Protein Unfolding and Receptor-Ligand Unbinding

In a simple picture, the unfolding of protein domains or unbinding of a receptor-ligand pair can typically be estimated as a thermally driven two-state system separated by a free energy barrier within a one-dimensional energy landscape, where applied force reduces the height of the energy barrier. The most noted theoretical framework to describe the unfolding behavior of such a system under force (and with moderate loading rates) is the Bell-Evans model [105–107].

Considering a constant force loading rate  $\dot{F}$  the Bell-Evans model predicts the force-dependent unfolding or unbinding probability

$$p(F) = \frac{k_{\text{off}}^0}{\dot{F}} \exp\left(\frac{F\Delta x}{k_{\text{B}}T} - k_{\text{off}}^0 \frac{k_{\text{B}}T}{\dot{F}\Delta x} \left(\exp\left(\frac{F\Delta x}{k_{\text{B}}T}\right) - 1\right)\right) \quad (2.3)$$

with force  $F$ , zero-force off rate  $k_{\text{off}}^0$ , distance from the bound to the transition state  $\Delta x$  and temperature  $T$ .

Differentiating equation 2.3 yields the most probable unfolding force  $F^*$ :

$$F^* = \frac{k_{\text{B}}T}{\Delta x} \ln\left(\frac{\dot{F}\Delta x}{k_{\text{off}}^0 k_{\text{B}}T}\right) \quad (2.4)$$

This linear relation between the logarithm of the force loading rate  $\dot{F}$  and the most probable unfolding force  $F^*$  is regularly used to extract the unfolding parameters  $k_{\text{off}}^0$  and  $\Delta x$  from so-called dynamic force spectra, where the pulling speed or loading rate is typically varied over a range of several orders of magnitude.

## 2.2 AFM Cantilever Calibration

In order to calculate actual force units from the measured photodiode deflection volts, two values have to be obtained from an AFM and its cantilever: The spring constant of the cantilever  $k$  and the inverse optical lever sensitivity (InvOLS).

A great variety of methods have been developed to obtain these parameters or calibrate the stiffness of an AFM cantilever in some other way, though only few fulfill the necessary conditions to be practical when applied in AFM-based SMFS [108]. Most importantly, the calibration should be possible *in situ*, thus use the AFM itself and not require any additional equipment, and be non-destructive for the tip of the lever, in order to keep any fictionalization intact.

The thermal vibration method is most widely used and relies on the assumption that cantilever acts like a harmonic oscillator and will fluctuate in response to thermal noise [109]. The Hamiltonian of a harmonic oscillator is given by

$$H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2 \quad (2.5)$$

where  $p$  is the momentum,  $m$  is the mass,  $\omega$  is the resonance frequency and  $q$  is the displacement of the oscillator. Following the equipartition theorem, each degree of freedom - thus the average value of each quadratic term in the Hamiltonian - is given by  $\frac{k_{\text{B}}T}{2}$ , which leads to

$$\langle \frac{1}{2} m \omega^2 q^2 \rangle = \frac{k_B T}{2} \quad (2.6)$$

Since the resonance frequency of a harmonic spring is given by  $\omega = \sqrt{k/m}$ , solving equation 2.6 for  $k$  results in:

$$k_{\text{cantilever}} = \frac{k_B T}{\langle q^2 \rangle} \quad (2.7)$$

The spring constant of a cantilever can thus be calculated from its mean square spring displacement. For a typical cantilever with a spring constant of 100 pN/nm thermal fluctuations will be of the order of 2 Å, which is about the size of a water molecule.  $\langle q^2 \rangle$  can therefore be approximated by the mean square displacement of the cantilever's fundamental vibrational mode  $z_1^2$ , since higher modes are not expected to have a major contribution at such small thermal fluctuations:

$$k_{\text{cantilever}} = \frac{k_B T}{z_1^2} \quad (2.8)$$

This equation was later modified by Butt and Jaschke, introducing a correction factor for the shape and deflection sensitivity of a rectangular cantilever when using only the fundamental resonance bending mode [110]:

$$k_{\text{cantilever}} = 0.971 \cdot \frac{k_B T}{z_1^2} \quad (2.9)$$

The InvOLS can be obtained by pushing the tip of the cantilever just hard enough into the sample surface to see a linear response in the resulting deflection volts *vs.* z-distance plot. Fitting a line to this response will yield a slope in the units V/nm, thus the change in deflection volts per nanometer bending at the tip of the lever. The InvOLS is the inverse of this slope and can therefore be used to calculate the change in z-position of the tip from deflection volts.

The resulting calculation from raw deflection volts to force units is therefore given by

$$\text{Force[pN]} = k[\text{pN/nm}] \cdot \text{InvOLS}[\text{nm/V}] \cdot \text{Deflection[V]} \quad (2.10)$$

The overall force uncertainty for this method was estimated to be at 15-20% [111]. Using a micro-electro-mechanical system (MEMS) and contact based method Brand *et al.* recently determined a deviation of 7.7% with a standard deviation of 15.1%. The main reason for this high deviation could be the high sensitivity of the spring constant from the position of the laser spot on the cantilever, which has been shown to change the measured stiffness up to 50% [112].

Multiplexed AFM-based SMFS measurements of several constructs with a single cantilever are thus not only beneficial to experimental throughput, but achieve excellent relative force comparability between the tested constructs, since cantilever calibration errors can be circumvented.





**Part III**  
**Results**



# 3

## Peer-Reviewed Research Articles

### 3.1 PUBLICATION P1: Combining *in Vitro* and *in Silico* Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

This article gives a detailed description of the mechanical properties of all individual cohesin domains from the cellulosomal scaffoldin ScaA of *A. cellulolyticus*. The mechanical stability of each domain within a cellulosomal scaffoldin has never been investigated or published before. Previously, only few selected domains of other scaffoldins and organisms had been studied by AFM-based SMFS [11].

The AFM-based SMFS assay used for this study establishes a new level of multiplexed force spectroscopy. A fast sample preparation protocol for the one-pot *in vitro* expression and surface attachment of multiple proteins in parallel was developed. This way it was possible to continuously probe all seven of ScaA's cohesins using a single cantilever, resulting in excellent relative force comparability compared to traditional methods. We found that the bridging cohesins, which are subjected to lateral forces in *A. cellulolyticus*' cellulosome, exhibit notably higher mechanical stabilities than the hanging cohesins. This finding is in agreement with the aforementioned limited studies and supports the conclusion that this trait may have been adapted to keep the cellulosome intact and working in challenging environmental conditions.

Furthermore, protein homology models were obtained for all cohesin domains and unfolded multiple times *in silico*. Force propagation pathways and correlation communities were calculated from the unfolding traces and this information was used to identify regions responsible for the surprisingly low unfolding forces of ScaA's mechanically weakest cohesin domain, cohesin 1. We were able to propose mutants to increase the mechanical stability of cohesin 1 by indirectly stabilizing its main structural element, the mechanical clamp motif. We compared the proposed mutants to the wild-type cohesins in a second AFM-based SMFS experiment and discovered that in one case a simple alanine to glycine point mutation suffices to more than double the force required to unfold the weakest of the ScaA cohesins. Thus, this work demonstrates that, even in the absence of high resolution structural protein models, steered molecular dynamics simulations provide valuable insights on a molecular level.

The stabilizing effect of dockerin binding to the cohesins mechanostability is presented in publication P2 on page 69.

# Combining *in Vitro* and *in Silico* Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

Tobias Verdorfer, Rafael C. Bernardi, Aylin Meinhold, Wolfgang Ott,  
Zaida Luthey-Schulten, Michael A. Nash and Hermann E. Gaub

*published in*

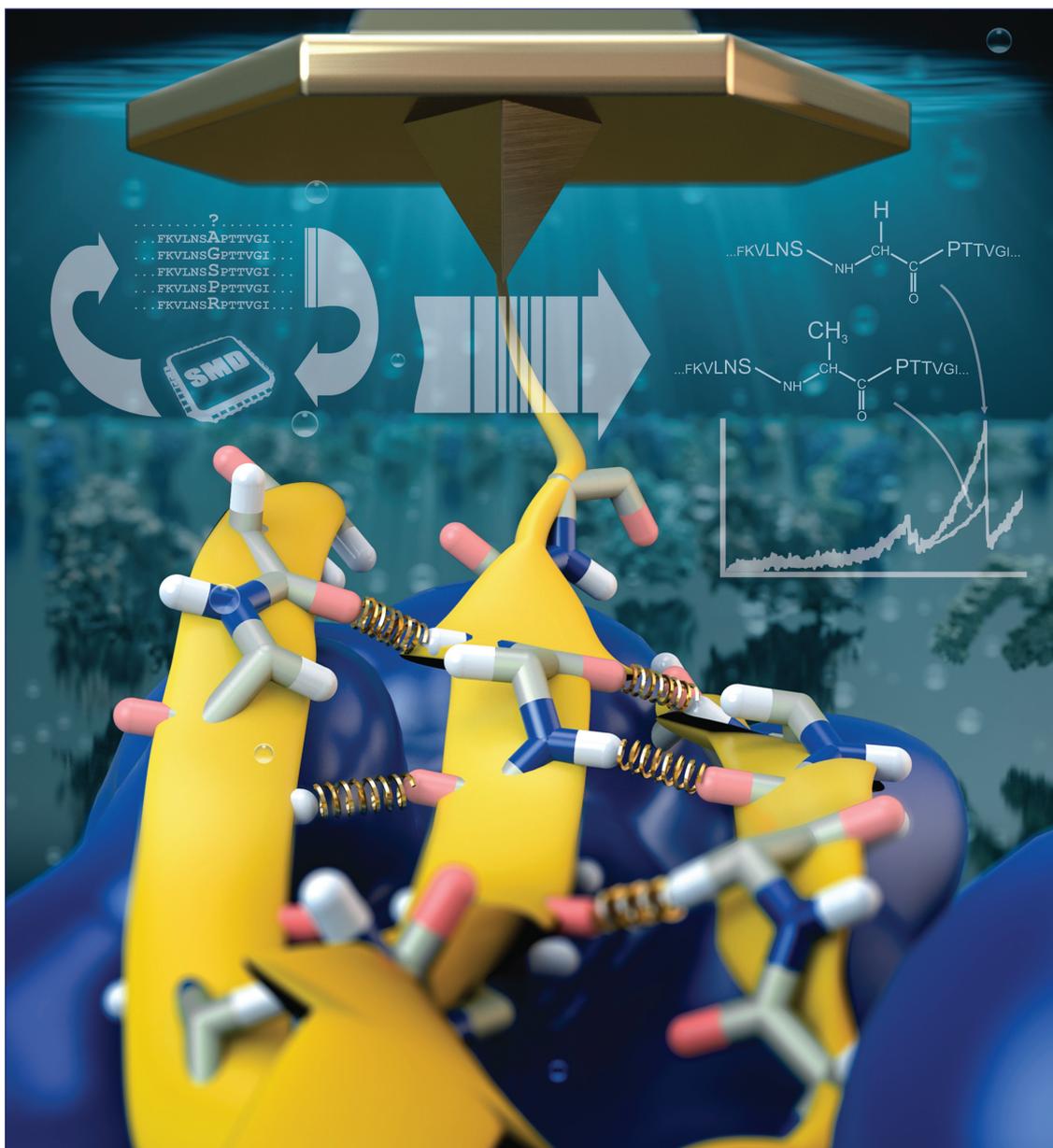
Journal of the American Chemical Society, 139(49), 17841-17852, (2017)

*Reprinted from [113], with permission from American Chemical Society  
Copyright 2017, American Chemical Society*

December 13, 2017  
Volume 139  
Number 49  
pubs.acs.org/JACS

# J | A | C | S

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY



 ACS Publications  
Most Trusted. Most Cited. Most Read.

[www.acs.org](http://www.acs.org)

## Combining *in Vitro* and *in Silico* Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

Tobias Verdorfer,<sup>†</sup> Rafael C. Bernardi,<sup>‡</sup> Aylin Meinhold,<sup>†</sup> Wolfgang Ott,<sup>†</sup> Zaida Luthey-Schulten,<sup>‡,§</sup> Michael A. Nash,<sup>\*,||,⊥</sup> and Hermann E. Gaub<sup>†,⊞</sup>

<sup>†</sup>Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

<sup>‡</sup>Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

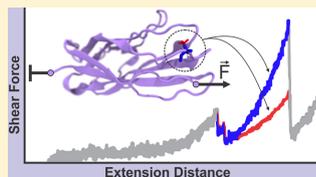
<sup>§</sup>Department of Chemistry, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801, United States

<sup>||</sup>Department of Chemistry, University of Basel, 4056 Basel, Switzerland

<sup>⊥</sup>Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland

### Supporting Information

**ABSTRACT:** Cellulosomes are polyprotein machineries that efficiently degrade cellulosic material. Crucial to their function are scaffolds consisting of highly homologous cohesin domains, which serve a dual role by coordinating a multiplicity of enzymes as well as anchoring the microbe to its substrate. Here we combined two approaches to elucidate the mechanical properties of the main scaffold ScaA of *Acetivibrio cellulolyticus*. A newly developed parallelized one-pot *in vitro* transcription–translation and protein pull-down protocol enabled high-throughput atomic force microscopy (AFM)-based single-molecule force spectroscopy (SMFS) measurements of all cohesins from ScaA with a single cantilever, thus promising improved relative force comparability. Albeit very similar in sequence, the hanging cohesins showed considerably lower unfolding forces than the bridging cohesins, which are subjected to force when the microbe is anchored to its substrate. Additionally, all-atom steered molecular dynamics (SMD) simulations on homology models offered insight into the process of cohesin unfolding under force. Based on the differences among the individual force propagation pathways and their associated correlation communities, we designed mutants to tune the mechanical stability of the weakest hanging cohesin. The proposed mutants were tested in a second high-throughput AFM SMFS experiment revealing that in one case a single alanine to glycine point mutation suffices to more than double the mechanical stability. In summary, we have successfully characterized the force induced unfolding behavior of all cohesins from the scaffoldin ScaA, as well as revealed how small changes in sequence can have large effects on force resilience in cohesin domains. Our strategy provides an efficient way to test and improve the mechanical integrity of protein domains in general.



### INTRODUCTION

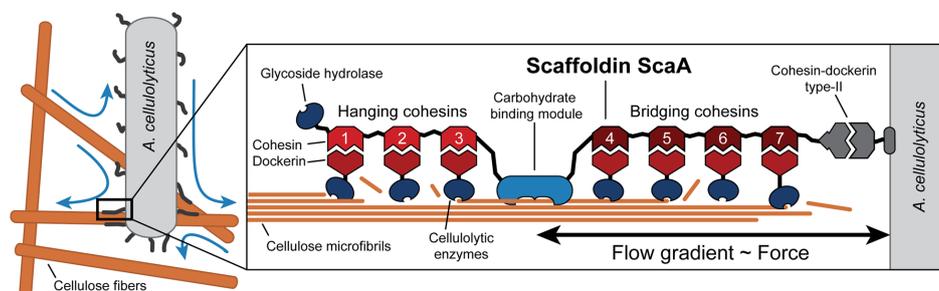
Multidomain protein scaffolds organize cellulolytic enzymes and provide adhesion between the host cell and its substrate. In cellulosomes, these so-called scaffoldins utilize various orthogonal high-affinity receptor–ligand interactions between cohesins and dockerins to anchor themselves to the cell's exterior membrane and to coordinate a broad arsenal of cellulolytic enzymes.<sup>1–3</sup> Cellulosomes are extracellular polyprotein complexes produced by many microorganisms for the efficient degradation of cellulose and hemicellulose, two of nature's most abundant polymers. In addition to multiple repeats of cohesin domains, scaffoldins may also contain other ancillary domains, such as enzymatic subunits or carbohydrate-binding modules (CBMs).<sup>4</sup> Cellulosomes represent an advantageous strategy compared to the secretion of freely soluble enzymes for cellulose degradation, because they achieve both robust adhesion to the substrate and synergistic and cooperative interplay between the enzymes. This complex synergy is based on the spatial organization and enhanced local

concentration. Due to their remarkable ability to achieve these complex tasks, cellulosomes have become a prime instructive example of molecular systems offering modularity, self-assembly, and highly efficient enzymatic catalysis.<sup>5</sup> Aside from the unique role they play in the degradation of cellulolytic material, cellulosomal scaffoldins are especially interesting as building blocks in a biotechnological toolbox. Cohesin–dockerin interactions with orthogonal specificities can be used to post-translationally incorporate selected enzymes or other auxiliary domains in specific locations by self-assembly within rationally designed multicomponent complexes.<sup>6–8</sup>

In this study, we focused on a scaffoldin of the cellulosome-producing organism *Acetivibrio cellulolyticus*, a gram-negative, anaerobic bacterium found in mechanically demanding environments, such as sewage sludge<sup>9</sup> or the bovine rumen.<sup>10</sup> Although *A. cellulolyticus* expresses an exceptionally branched and diverse

Received: July 28, 2017

Published: October 23, 2017



**Figure 1.** Simplified schematics of the cellulosome of *A. cellulolyticus*. The scaffoldin ScaA binds up to 7 cellulolytic enzymes via cohesin–dockerin interactions and attaches the multienzyme complex to the cellulose microfibril through a CBM module, ensuring close proximity of the enzymes and substrate. Simultaneously, ScaA anchors the assembled cellulosome to the cell wall of the microbe via a type-II cohesin–dockerin complex. The bridging cohesins 4–7 in particular have to withstand the forces between the cellulose fibril and microbe caused by flow gradients in the environment.

cellulosome,<sup>11</sup> the assembly of the majority of its cellulolytic enzymes relies mainly on a single scaffoldin (ScaA)<sup>12</sup> (Figure 1). Simultaneously, ScaA has the essential function of targeting specific substrates and anchoring the cell to the cellulose fibril via its single CBM. ScaA therefore serves a dual purpose for the organism by incorporating catalytic modules into the cellulosome complex and by adhering the bacterium to the cellulose substrate, making it the key player in *A. cellulolyticus* cellulosome.

ScaA consists of an N-terminal glycoside hydrolase domain, seven type-I cohesins, a CBM located between cohesins 3 and 4, and a C-terminal type-II dockerin domain. Cohesins 4 through 7 are located between the CBM and the anchoring type-II Doc and form the so-called “bridging” region of the scaffoldin, while cohesins 1 through 3 are located exterior from the CBM and form the “hanging” region of this scaffoldin. It had been hypothesized that the cohesins of the bridging region will be subjected to higher mechanical stress compared to the hanging cohesins because the CBM anchors the cell to the cellulose substrate and the domains located within the hanging region are not expected to be mechanically stressed *in vivo*.<sup>13</sup> The organism may have adapted the bridging cohesins to these conditions, and therefore, they should be able to withstand higher mechanical stress in order to maintain a folded structure.

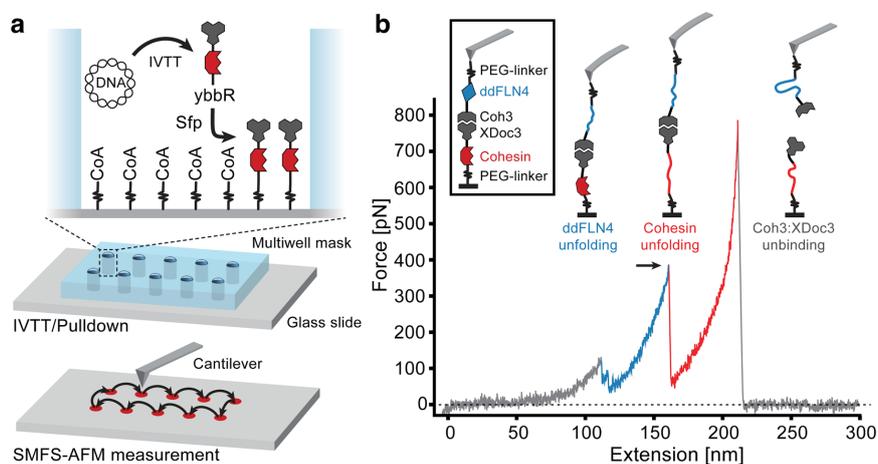
This hypothesis of bridging versus hanging cohesins within cellulosomal scaffoldins has previously been successfully tested in part for the CipA scaffoldin of *Clostridium thermocellum*. These prior results showed that cohesins from the bridging region withstood higher unfolding forces compared to those from the hanging region.<sup>13</sup> This behavior has been attributed to slight differences in the stability of the mechanical clamp motifs, which are structural elements formed by backbone hydrogen bonds between parallel  $\beta$ -strands of the N- and C-terminal ends of the cohesin protein domains.

Since the mechanobiology of cellulosomes is key to their function, we investigated the mechanostability of cohesins of the ScaA scaffoldin as well as several derived mutants at the single molecular level. The high sequence similarity between the selected cohesins raises the question of how differently these cohesins withstand mechanical stress. It is known that small variations in cohesin primary sequences are responsible for determining the specificity of interactions with their dockerin binding partners,<sup>14,15</sup> but very little is known about

how sequence variations affect mechanical stability. To address these points, we performed automated atomic force microscope (AFM)-based single molecule force spectroscopy (SMFS) experiments, a technique that has been established as a robust standard approach to investigate the mechanics of individual molecules.<sup>16–22</sup> The fast dynamics and particularly the large force range of AFM-SMFS made it a prime choice for our investigations. To ensure improved relative force comparability and high experimental throughput, we developed a parallelized AFM sample preparation method, which utilizes a one-step protein expression and surface immobilization protocol which is a simplified and easy-to-use version of work previously introduced by our group.<sup>23</sup> At the same time, we carried out steered molecular dynamics (SMD) simulations on structural protein models derived from a homology modeling strategy, which improved our understanding of the cohesin mechanostability at the submolecular level. By employing cross-correlation based network analysis on simulation trajectories, we identified regions involved in structural stability outside the mechanical clamp motif and proposed mutations to pin down single crucial amino acids involved in fold strength. This so-called dynamical network analysis has been successfully employed before to investigate force propagation pathways in cohesin–dockerin complexes<sup>24</sup> and in filamins.<sup>25</sup> We tested the proposed mutants experimentally to verify the influence of the proposed alterations on mechanical stability. Employing this combined computational/experimental approach, we were able to predict and verify, among others, a single point mutation outside of the mechanical clamp motif of cohesin 1, which increased its mechanical stability 2.6-fold. The ability to predict such a remarkable difference in mechanostability reveals the potential of our combined approach to characterize and manipulate the mechanical properties of protein domains.

## RESULTS

**AFM-Based SMFS.** As specific pulling handles throughout this study we used a type-III cohesin–dockerin complex (Coh3:XDoc3) from *Ruminococcus flavifaciens*, which ensured a high yield of single-molecule interaction curves because of its high specificity and long-term stability. This pulling handle enabled the unfolding of mechanically rigid cohesins due to its high receptor–ligand rupture force of more than 600 pN.<sup>26</sup> It is important to note here that the Coh3:XDoc3 interaction, which



**Figure 2.** Schematics of the experimental design and exemplary force curve. (a) A multiwell mask is attached to a glass slide and the surface is functionalized with PEG–Coenzyme A. In a one-pot reaction, an IVTT-kit expresses the proteins of interest containing a ybbR-tag at their N-termini, and an XDoc3 domain at their C-termini. Sfp catalyzes a reaction to bind the constructs site-specifically and covalently to the PEG–CoA spacers. After removal of the multiwell mask, the different constructs are probed by the same AFM tip in an automated fashion. (b) Exemplary force–distance trace showing the unfolding of ybbR–cohesin–XDoc3 immobilized on the surface and the Coh3–ddFLN4–ybbR bound to the cantilever, when the cantilever is retracted from the surface with constant speed. A typical curve shows a two-step unfolding and subsequent stretching of the ddFLN4 fingerprint domain (blue), followed by the unfolding (indicated by an arrow) and stretching of the cohesin domain under investigation (red) and a final rupture of the Coh3:XDoc3 complex.

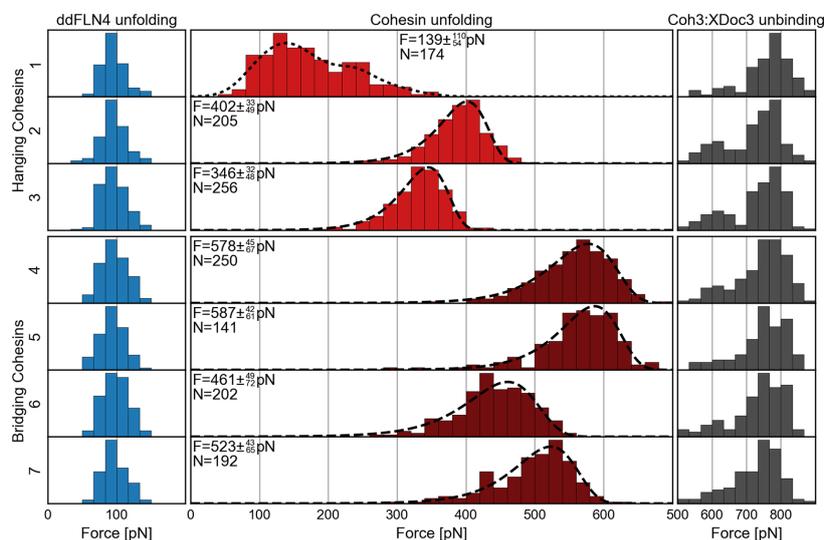
we employed as a specific pulling handle, is orthogonal to the ScaA cohesin domains that were unfolded under force and thus does not interact with them. ScaA cohesin domains from the bridging or hanging region were cloned in frame with XDoc3 and synthesized using cell-free expression (see below). The cohesins of interest therefore carried the XDoc3 domain at their C-termini, which was able to bind to the Coh3 on the cantilever. The cantilever carried covalently and site-specifically immobilized Coh3 domains, which were recombinantly expressed as fusion proteins with the fourth immunoglobulin rod filamin domain from *Dictyostelium discoideum* (ddFLN4).<sup>27,28</sup> We used the ddFLN4 as fingerprint domain within the molecular chain to reliably rule out data traces with unspecific or multiple interactions. The fingerprint domain unfolds at relatively low forces ( $\sim 100$  pN) and its contour length increment ( $\sim 34$  nm) differs significantly from the expected contour length increments of the cohesins ( $\sim 45$  nm). Furthermore, based on our experience we knew that ddFLN4 reliably refolds following mechanical unfolding when bound to the cantilever.

As calibration errors of up to 15% come with cantilever-based force measurements,<sup>29</sup> the precision and therefore the comparability of individual SMFS experiments is limited by the accuracy of the calibration of the cantilever spring constant. This limitation can be circumvented by presenting a set of molecules in a covalently linked microarray format on a single glass slide in predetermined positions.<sup>23</sup> Taking advantage of this setup, the proteins of interest can all be measured with a single cantilever in a high-throughput fashion, and large sets of molecular constructs can be compared in a single AFM measurement session. Although the systematic error of cantilever calibration will still be up to 15%, the relative stability of the proteins can be assessed with extremely high

precision, allowing us to detect differences in stability well below 15%.

We have previously reported the use of microfluidics in combination with an intricate multistep *in vitro* on-chip protein expression and covalent surface attachment protocol.<sup>23</sup> Here we were able to reduce the complexity of the setup significantly, gaining throughput as well as flexibility. We developed a new sample preparation protocol, which is schematized in Figure 2a. Briefly, a silicone mask was placed on a glass slide to form microwells, and the resulting spots were covalently functionalized with PEG–Coenzyme A. The individual wells were filled with an *in vitro* transcription/translation (IVTT) system, along with the plasmid DNA encoding for the fusion proteins and phosphopantetheinyl transferase (Sfp).<sup>30</sup> In a single incubation step, this mixture resulted in cell-free protein synthesis, and simultaneous covalent ligation of the protein library onto the surface through a ybbR-tag<sup>30</sup> at the N-terminus. The mask was removed from the glass slide and the spatially separated protein spots were probed in series using a single functionalized cantilever. The combination of a site-specific N-terminal enzyme-mediated immobilization strategy, and a specific C-terminal pulling handle ensured that only fully expressed constructs were probed by the AFM. We recorded 2000 pulling cycles per protein spot while continuously cycling through the array in an automated fashion. No further user interaction was required after the start of the experiment, which allowed long-term multiday measurements to build up large statistics for each construct.

A typical force versus distance trace is shown in Figure 2b. The cantilever approached the surface, and Coh3 bound to XDoc3. Upon retraction of the cantilever with constant speed, the polypeptide chain stretched until the ddFLN4 fingerprint unfolded in a distinct two-step pattern, followed by the



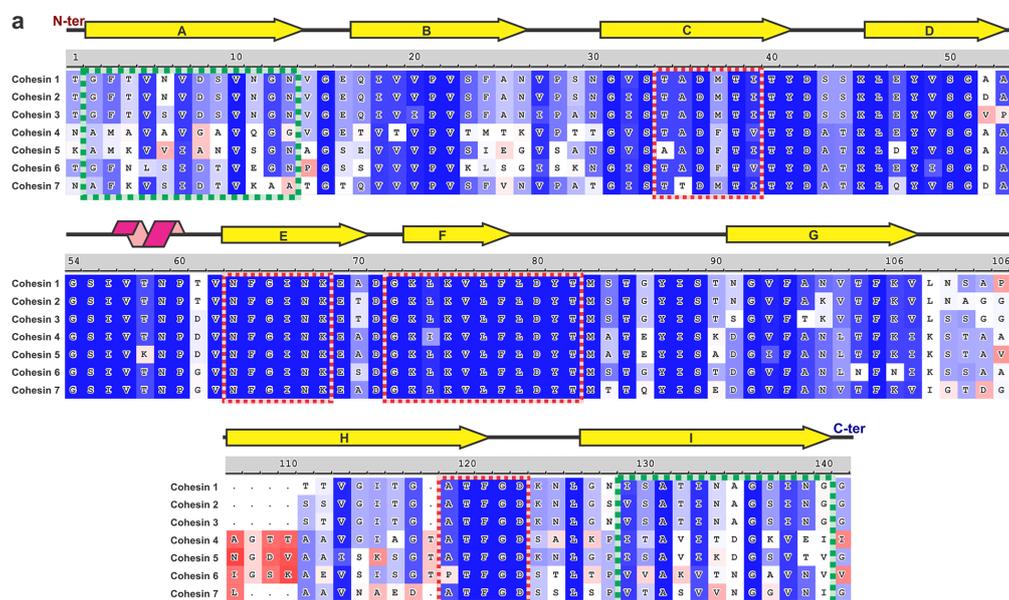
**Figure 3.** Unfolding and rupture force histograms of wild-type ScaA cohesins. Histograms showing the unfolding and unbinding forces of the fingerprint domain ddFLN4 (blue), the hanging (light red) and bridging (dark red) cohesins of ScaA, and the Coh3:XDoc3 receptor–ligand pulling handle (gray). The force distributions of the ddFLN4 fingerprint and the Coh3:XDoc3 handle are independent of the measured ScaA cohesin construct, which allows for improved relative force comparability of the ScaA cohesins. Unfolding force distributions of cohesins 2–7 were fitted following the Bell–Evans model (dashed lines). A kernel density estimation (KDE) was used to obtain the most probable unfolding force ( $\pm$ fwhm) of cohesin 1 (dotted line). All data were recorded using a single cantilever with a spring constant of 225 pN/nm at a retraction speed of 1600 nm/s during a 24 h automated SMFS experiment.

unfolding of the respective cohesin under investigation. Finally the Coh3:XDoc3 bond ruptured typically at forces of  $\sim$ 780 pN at force loading rates of  $\sim$ 10<sup>5</sup> to 10<sup>6</sup> pN/s. The heights of the various peaks are plotted in the histograms of Figures 3 and 5. Only force curves displaying a distinct combination of contour length increments (ddFLN4  $\sim$  34 nm,<sup>27,28</sup> cohesin  $\sim$  45 nm) were included in our analysis. Alignments of all relative contour length increments from all pulling experiments can be found in the Supporting Information (Supplemental Figure S1).

**AFM-Based SMFS on ScaA's Wild-Type Cohesins.** To investigate the mechanical stability of *A. cellulolyticus*' scaffoldin ScaA, seven distinct surface spots (one for each cohesin) were prepared using the one-step expression/immobilization reaction described above. All data were collected using a single cantilever. Figure 3 shows the resulting unfolding and rupture force distributions. The outer histograms show very similar force distributions of the fingerprint domain ddFLN4 and of the pulling handle Coh3:XDoc3 independent of the measured molecular construct, which agreed with previously reported literature values.<sup>26,27</sup> Coh3:XDoc3 showed a sharp peak at  $\sim$ 780 pN and a minor shoulder at  $\sim$ 600 pN, both of which are known features of this molecular complex.<sup>26</sup> The remarkable consistency in force among different molecular constructs indicated low force drift during the course of the experiment and confirmed our ability to achieve precise relative comparability of the unfolding forces of the cohesins. Following work done by the Schulten<sup>31</sup> group and Evans and Ritchie<sup>32</sup> in 1997, the cohesin unfolding force distributions can be fitted using a two-state model (from here on called the Bell–Evans model).

With the exception of cohesin 1, all unfolding force distributions could be fitted using the Bell–Evans model, indicating a stable fold and a single barrier unfolding pathway dominating the unfolding process, and most probable rupture forces were obtained. The measured unfolding forces of cohesin 1 resulted in a more complex force distribution. Individual AFM force–distance traces gave no hints of any peculiarities in comparison to the unfolding traces of the other cohesins. We hypothesize that cohesin 1 exhibits more than one distinct fold or has several multibarrier unfolding pathways that precluded its unfolding force distribution to be fitted using a simple two-state model. In this case, we used kernel density estimation (KDE) as a means to smooth the unfolding force histogram and obtain the most probable unfolding force. As originally hypothesized, the three hanging cohesins showed in fact a considerably lower most probable unfolding force (for full width at half-maximum (fwhm) errors, see Figure 3) (cohesin 1, 139 pN; cohesin 2, 402 pN; cohesin 3, 346 pN) compared to the four bridging cohesins (cohesin 4, 578 pN; cohesin 5, 587 pN; cohesin 6, 461 pN; cohesin 7, 523 pN).

**Cohesin Homology Models.** Since structural data were not available for any of the ScaA cohesins, a homology modeling strategy was adopted,<sup>33</sup> employing Modeller 9.17<sup>34</sup> to obtain structural models for all the cohesins investigated here (Figure 4b). Using BLAST,<sup>35</sup> we obtained homologous cohesin structures (PDB IDs 1G1K, 4DH2, 2VN6, and 4UMS) within the Protein Data Bank<sup>36</sup> (PDB). These structures were then used as templates to derive the homology models that were further refined with molecular dynamics (MD) simulations. Equilibration for 100 ns was performed using NAMD<sup>37</sup>

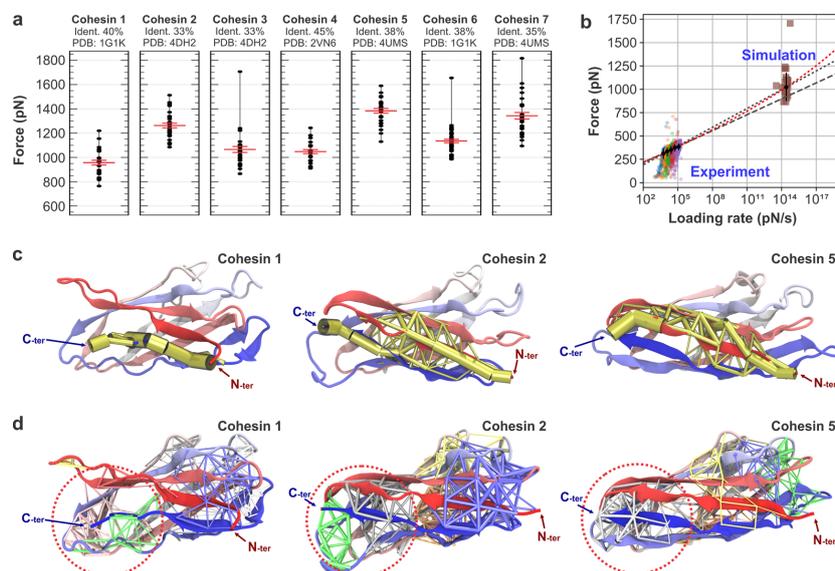


**Figure 4.** Sequence alignment and modeling workflow. (a) Sequence alignment of all cohesins of ScaA reveals high overall conservation. The amino acids thought to be primarily involved in mechanical stability<sup>13</sup> are represented by green boxes. Regions primarily involved in dockerin recognition and binding are represented in red boxes. The background colors of the letters represent BLOSUM 70 sequence alignment score, from high (dark blue) to low (red). A high resolution version can be found in Supplemental Figure S6. (b) Schematic representation of molecular modeling and dynamics protocol. Homologous protein structures were obtained by running the sequences of the cohesins against the Protein Data Bank using BLAST. Most highly identical structures with high sequence coverage were then used as templates to obtain structural models using Modeller 9.17. Models were properly solvated employing QwikMD and equilibrated for 100 ns of unbiased molecular dynamics simulations using NAMD. A very similar fold was observed for all seven cohesins, here colored ranging from red to blue based on its residue index number. SMD simulations were performed by holding the N-termini and pulling the C-termini with constant velocity.

through its QwikMD interface<sup>38</sup> (see Supplemental Figures S2, S3, and S4). Comparison of the aligned structures (Supplemental Figure S5A) reveals that all seven cohesins of ScaA show a similar general fold. Even though the range of identity between our model cohesins and the best available structural templates was between 33% and 45%, cohesins always present a very similar fold, helping the structure prediction.

**SMD Simulations and Network Analyses.** To evaluate the behavior of ScaA's cohesins under force, we performed SMD simulations<sup>31,39</sup> using NAMD and QwikMD, where the N-termini of the constructs were fixed and the C-termini were pulled with constant velocity. Employing four different pulling speeds, we investigated first all the unfolding steps in long (on the order of microseconds) SMD simulations. For all cohesins, the first part to unfold with highest peak force is the C-terminal region with  $\beta$ -strand I, followed by  $\beta$ -strand H losing its structure (see Supplemental Figure S7). Next, in the N-

terminal region, both  $\beta$ -strands A and B lose their structure almost at the same time in most simulations.  $\beta$ -strands C–G finally lose structure under relatively low force (see representative unfolding analysis in Supplemental Figure S8). The force necessary to break any of the cohesin folds was observed to be equal to the force required to unfold the C-terminal segment of these cohesins. Therefore, to obtain statistically relevant mechanical stability behavior, we performed many 20 ns long simulation replicas (at least 25 replicas per construct). Figure 5a shows the peak unfolding forces, as well as the sequence identity between the cohesins of ScaA and their respective best available structural template. With few exceptions, the simulation replicas revealed qualitatively the same general trend of the unfolding force peak distributions as measured by SMFS (Figure 3). The absolute forces however are shifted in the simulations toward higher values (Supplemental Figure S9), which was to be expected due to



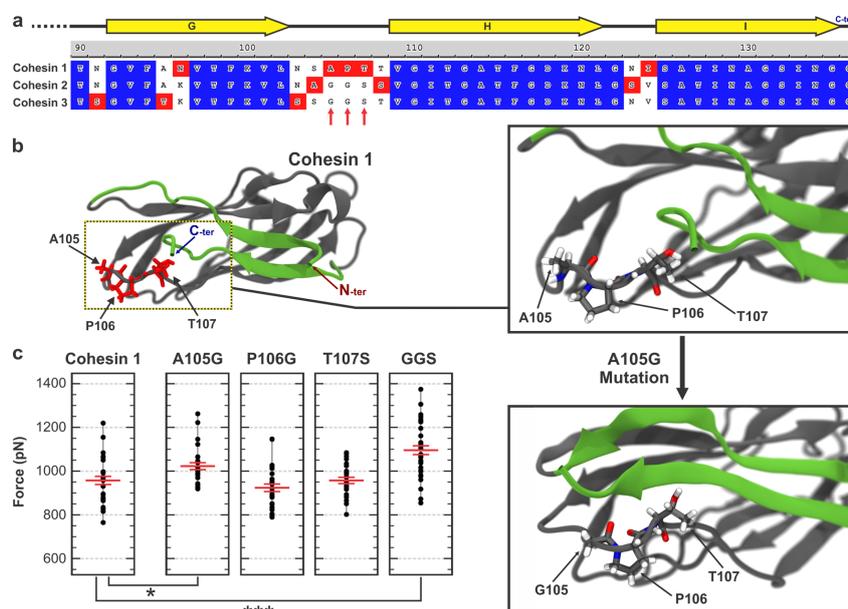
**Figure 5.** SMD peak force results, dynamic force spectrum, force propagation pathways, and community analysis. (a) SMD peak unfolding forces from each simulation replica (black dots) and average peak force per cohesin (red line  $\pm$  SEM) for the different natural cohesins of ScaA. Sample force versus distance profiles are found in Supplemental Figure S10. Statistical significance between the unfolding forces of all cohesins can be found in Supplemental Figure S11. Also shown is the sequence identity to the respective PDB homology modeling template. (b) Dynamic force spectrum for unfolding events of cohesin 3. Varicolored points represent rupture force/loading rate data from an experiment with 5 different pulling speeds. Brown squares represent rupture force/loading rate data from SMD simulations. Black points represent the most probable rupture force/loading rate of each pulling speed obtained from kernel density estimates. Error bars represent the fwhm. Gray lines represent least-squares fits of the Bell-Evans model to the experimental and to both the experimental and the simulation data with fitting parameters ( $\pm$ SD)  $\Delta x = 0.17 \pm 0.012$  nm,  $k_0 = (6.7 \pm 6.3) \times 10^{-4} \text{ s}^{-1}$  and  $\Delta x = 0.14 \pm 0.0015$  nm,  $k_0 = 4.9 \times 10^{-3} \pm 8.9 \times 10^{-4} \text{ s}^{-1}$ , respectively. The red dotted line represents a least-squares fit of the DHS model to the combined experimental and the simulation data with fitting parameters ( $\pm$ SD) of  $\Delta x = 0.19 \pm 0.024$  nm,  $k_0 = (1.4 \pm 2) \times 10^{-4} \text{ s}^{-1}$ , and  $\Delta G = 60 \pm 13 k_B T$ . For detailed representation of experimental data, see Supplemental Figure S12. (c) Force propagation pathways through selected cohesins calculated using Pearson correlation (yellow tubes). The thickness of the pathway edges represents the normalized probability of force propagating through the particular edge. (d) Network-based community analysis in selected cohesins showing regions with high internal correlation during pulling simulations calculated using generalized correlation. Communities are colored individually, and thick connections correspond to high correlation.

much higher force loading rates of the simulations compared to the AFM experiments.

A force offset between MD simulations and SMFS experiments can easily be understood in view of the Bell-Evans model, which predicts a linear dependence of the unfolding force from the logarithm of the force loading rate. To corroborate this assumption, we varied the pulling rate in the SMFS experiments of cohesin 3 and plotted the measured unfolding forces together with the corresponding MD simulations in Figure 5b. The dashed line represents the best fit to the experimental data extrapolated to the MD time domain; the dotted line fits both the experimental and the simulation data. The resulting distances to the transition state of  $0.14 \pm 0.012$  nm and  $0.17 \pm 0.0015$  nm agree very well with literature values of comparable cohesin protein structures.<sup>13,15,26</sup> It is important to note that the slope in the dynamic force spectrum can change with increasing pulling speeds, resulting in a nonlinear upturn at higher pulling velocities as shown by Rico et al.<sup>18</sup> This effect is caused by a shift from a stochastic to a deterministic unfolding regime. In the former, the unfolding process is governed by spontaneous,

thermal unfolding under a given force, while in the latter, the high pulling velocities leave the protein insufficient time to sample its energy landscape. As described in the Dudko, Hummer, and Szabo model (DHS model), the regime transition can happen at different loading rates and is characterized by the critical force,  $F_c = 2\Delta G/\Delta x$ .<sup>40</sup> The position of the transition from stochastic to deterministic regime therefore strongly depends on the general mechanical stability of the system under investigation. Fitting the DHS model to the data in Figure 5b resulted in a critical force  $F_c \approx 2500$  pN, suggesting that our SMD simulations were carried out at loading rates where unfolding is still dominated by stochastic fluctuations. This indicates that the SMD simulations provide an accurate description of the unfolding process of the system in this study.

The aforementioned results motivated a detailed analysis of the molecular structures and interactions, which could give rise to the particular properties of the different cohesins. Using Pearson cross-correlation-based force propagation analysis, a recently introduced protocol for the analysis of load distributions in molecular complexes,<sup>24</sup> we calculated the



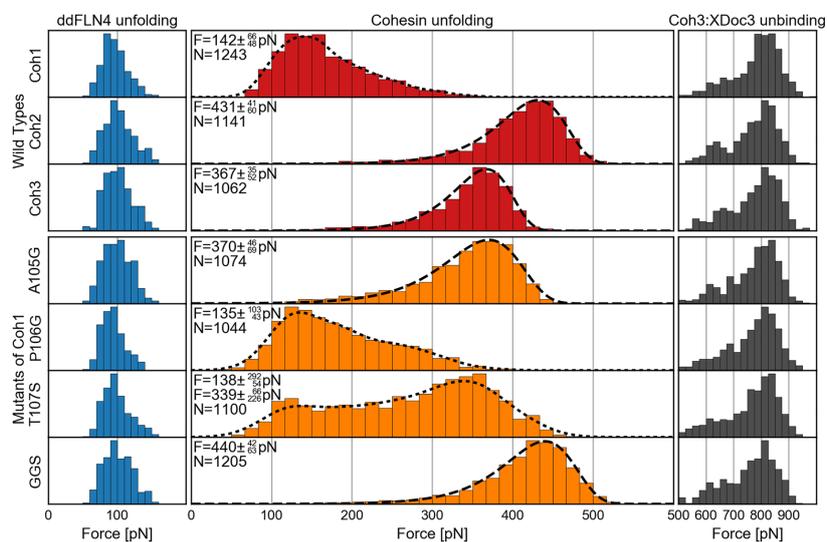
**Figure 6.** Mutant design and SMD peak force results. (a) Sequence alignment of the C-terminal end of the hanging cohesins 1–3. From the community analysis, we deduced that the low force resilience of cohesin 1 originated from the loss of sequence identity in the flexible area around amino acids 103–108. Based on comparison with cohesins 2 and 3, we designed three point mutations (A105G, P106G, and T107S) and a triple mutant (A105G P106G T107S (referred to as mutant “GGS”). A high resolution version can be found in [Supplemental Figure S15](#). (b) Left, homology model of cohesin 1 showing the  $\beta$ -sheets involved in the mechanical clamp motif in green and unique amino acids within the group of hanging cohesins in red. The linker region around amino acids A105, P106, and T107 was observed to be most flexible in the SMD simulations and was therefore suggested for mutation studies. Right, comparison of the homology models of cohesin 1 and its mutant A105G with molecular representation of the aforementioned region of interest. A seemingly small point mutation from an alanine to a glycine (from a methyl to a proton side chain) at position 105 changed the fold of the protein in this region significantly, resulting in much closer  $\beta$ -strands in the mechanical clamp motif. (c) SMD peak force from each simulation replica (black dot) and average peak force per cohesin (red line  $\pm$  SEM). The significantly increased peak forces of the mutants A105G and GGS suggest an increased stability compared to wild-type cohesin 1, while mutants P106G and T107S showed no significant change in average peak unfolding force. The single A105G mutation was able to recover forces in the same range of cohesin 3 (see [Figure 5a](#)). Statistical significance between the unfolding forces of all cohesins can be found in [Supplemental Figure S16](#).

suboptimal force propagation pathways, revealing that force propagates mainly through  $\beta$ -strands A, B, and I (see [Supplemental Figure S13](#)). [Figure 5c](#) shows the force propagation pathways through cohesin 1 and the two strongest cohesins within their respective group, cohesins 2 and 5. Cohesin 1 predominantly showed a single path between the N- and C-terminal  $\beta$ -sheets that carried the entire mechanical load, suggesting a badly formed mechanical clamp motif. The other cohesins show multiple possible force pathways suggesting a better distribution of force propagating from the N- to the C-terminus through a multitude of backbone hydrogen bonds (see [Supplemental Figure S13](#)).

In addition, we investigated the communities formed in the systems by employing dynamic network analysis<sup>41</sup> and generalized correlation<sup>42</sup> (see [Supplemental Figure S14](#)). These communities correspond to sets of residues that move in concert with each other and can be used to investigate regions that are generally more strongly connected during pulling simulations. [Figure 5d](#) shows the individual communities of cohesins 1, 2, and 5 in different colors, where thicker connections between the amino acids correspond to higher correlation between them. While cohesins 2 and 5 showed

pronounced communities connecting the C-terminus to its surroundings ([Figure 5d](#), red dashed circle region), cohesin 1 showed weak communities in this area of the protein, suggesting high and uncorrelated flexibility and therefore loose intraprotein contacts between the N- and C-terminal  $\beta$ -sheet. Particularly, the area between  $\beta$ -sheets G and H turned out to be most flexible in the case of cohesin 1. The same region of cohesin 1 was observed to be highly flexible also during the 100 ns MD equilibration, and was not as flexible in any other cohesin investigated here. Taken together, our results suggested that this is a critical region responsible for cohesin 1 lower stability under mechanical force.

**Mutant Design and SMD Simulations.** Since the region between  $\beta$ -sheets G and H in cohesin 1 was found to be the most flexible during the equilibrium MD and the analysis of cross-correlation communities suggested that weak communities in this same area could be responsible for the badly formed mechanical clamp motif of cohesin 1, we aligned the sequences of all hanging cohesins in this region ([Figure 6a](#)). Despite the high overall sequence similarity of  $\sim$ 85% among *A. cellulolyticus*'s hanging cohesins, major differences exist between the weaker cohesin 1 and the stronger cohesins 2 and 3 in the



**Figure 7.** Unfolding and rupture force histogram of the hanging cohesins and mutants of cohesin 1. Histograms showing the unfolding and unbinding forces of the fingerprint domain ddFLN4, the wild-type hanging cohesins 1–3 (red), mutants of cohesin 1 (orange), and the Coh3:XDdoc3 receptor ligand binding handle (gray). The force distributions of the ddFLN4 fingerprint and the Coh3:XDdoc3 handle are independent of the measured ScaA cohesin construct. Similar to Figure 3, force histograms were fitted following a Bell–Evans model where possible (cohesin 2, cohesin 3, A105G, and GGS; dashed line). A KDE was used to find the most probable rupture forces ( $\pm$ fw) in all other cases (cohesin 1, P106G, and T107S; dotted line). All data were recorded using a single cantilever with a spring constant of 163 pN/nm at a retraction speed of 1600 nm/s during a 72 h automated SMFS experiment.

region between  $\beta$ -strands G and H, more precisely from amino acid 103 to 108. Considering the results of the SMD simulations, force propagation, community analysis, and the sequence comparison with cohesin 2 and 3, we proposed four mutants of cohesin 1, namely, A105G, P106G, and T107S, and a triple mutant that we refer to here as GGS (A105G, P106G, and T107S). To investigate if these mutants would show a higher force resilience compared to their wild-type counterparts, we followed the same modeling and simulation steps as previously discussed for the wild-type cohesins.

After the 100 ns MD equilibration, mutated cohesins showed significant structural differences compared to the wild-type cohesin 1. A single alanine to glycine mutation (A105G), for example, already stabilized the nearby regions of the protein, resulting in a flawlessly folded  $\beta$ -stranded C-terminus, as shown in Figure 6b. SMD simulations, as shown in Figure 6c, revealed that A105G and GGS displayed a significantly higher unfolding force than cohesin 1, with A105G showing a mean unfolding force similar to that of cohesin 3. Analysis of the force propagation profiles and communities of the A105G mutant revealed a behavior that resembles one of the stronger cohesins, as shown in Supplemental Figures S13 and S14.

**AFM-Based SMFS on Hanging Cohesins and Mutants of Cohesin 1.** In order to test the predictions from the SMD simulations and to identify the amino acids responsible for the low force resilience of cohesin 1, we prepared a second set of AFM-based SMFS experiments similar to the ones described in Figure 3. We compared the mechanical stability of the proposed mutants A105G, P106G, T107S, and the triple mutant GGS with the three hanging cohesins, again using a

single cantilever to ensure improved relative force comparability (Figure 7). Similar to Figure 3, force distributions not following the Bell–Evans model (cohesin 1, P106G, and T107S) were smoothed using a KDE to obtain meaningful most probable rupture forces. Most remarkably, mutant A105G showed a dramatic increase in most probable rupture force to 370 pN, making it around 2.6 times stronger than its parent structure cohesin 1 and therefore about as strong as cohesin 3. While the mutant unfolding simulations predicted a slight decrease in mechanical stability for the P106G mutant, the experimentally obtained unfolding force histogram shows no considerable change compared to wild-type. Mutant T107S exhibited a bimodal unfolding force distribution with most probable rupture forces of 138 pN and 339 pN roughly similar to the most probable unfolding forces of wild-type cohesins 1 and 3. A detailed examination of individual unfolding traces from different force regimes showed no distinctive features that could explain its bimodal unfolding force distribution. We can only theorize that this construct might exhibit a combination of strongly differing folded conformations or unfolding pathways. Such behavior was not observed in the simulations. The triple mutant GGS showed a most probable rupture force of 440 pN, making it as strong as cohesin 2.

## DISCUSSION

**Mechanical Stability of Highly Homologous ScaA Cohesins.** The high precision comparison of the mechanical stability of seven homologous cohesin domains from *A. cellulolyticus*' scaffoldin ScaA was enabled by the development of a novel SMFS sample preparation method, where several

constructs were produced by cell-free *in vitro* expression and covalent linkage to the surface in parallel in individual microwells. The proteins were probed sequentially with a single cantilever, enabling precise comparison of unfolding force distributions between multiple domains by eliminating relative error in calibration of cantilever spring constant values. All constructs contained a ddFLN4 domain, which served as a molecular fingerprint and allowed clear identification of single-molecule force–distance traces showing specific interactions through their unique contour length increments. This overall approach facilitated high-throughput SMFS of multiple proteins and allowed for fast and automated data analysis.

A clear trend can be seen in the most probable unfolding force of all cohesin domains from ScaA (Figure 3). The three cohesins on the far side of the substrate-anchoring CBM (i.e., hanging cohesins) showed in fact a considerably lower most probable unfolding force compared to the four bridging cohesins, which are exposed to force *in vivo*. This result strongly supports the hypothesis that higher mechanical stability is a repeatable feature of cohesins in the bridging region of cellulosomal scaffoldins, despite the fact that all domains tested show high levels of sequence homology (see Figure 4a).

To elucidate the origins of the large differences in mechanostability of the cohesin domains, we generated structural homology models for all seven cohesins under investigation. The combination of knowledge-based information from structural templates together with modern force fields and molecular dynamics allowed us to employ a real-space structural prediction and refinement strategy to obtain all ScaA cohesin structures.<sup>33</sup> However, the computational prediction of three-dimensional protein structures has its limitations, and the accuracy of the predicted models is strongly dictated by the availability of close structural templates.<sup>43</sup> The range of identity between our model cohesins and the best available structural templates was between 33% and 45%, implying rather poor homologies. To check the fold stability, all structural models were subjected to 100 ns of equilibrium MD, and the final structures were superimposed. The results showed that, even though four different templates were employed, all cohesins generated highly similar structural models (see Figure 4b and Supplemental Figure S5).

The seven structural models were then stretched *in silico* using a constant velocity SMD protocol. It is noteworthy that, with the exception of cohesin 4, the forces of all distributions were shifted by a constant value ( $\pm$ SD) of  $782 \pm 29$  pN (see Supplemental Figure S9). This finding is remarkable if one takes into account the relatively low identity between the modeled systems and their templates (33%–45%). Simulations showed that although cohesin 4 has the highest identity to its template (45%, PDB 2VN6), it might have been a suboptimal choice resulting in a nonideal folding state, as it shows an N-terminal region with wobbly  $\beta$ -strand formation (see Supplemental Figures S4 and S5).

Comparing the force peaks between the simulations and experiments served as a validation for the homology structures. Figure 5b shows a direct comparison between results obtained with AFM SMFS across a range of loading rates from  $\sim 10^3$  pN/s and  $\sim 10^5$  pN/s, and those obtained from *in silico* SMFS at  $\sim 10^{14}$  pN/s. Fitting the DHS model to the data suggests that the loading rates used in our SMD simulations fall into the stochastic regime. This finding, which bridges 11 orders of magnitude in force loading rate, indicates that the homology

models provide an accurate description of the unfolding process, validating the predictive power of both comparative modeling and *in silico* SMFS. We want to emphasize that this only holds true for remarkably strong proteins like the cohesins investigated here. At similar loading rates of  $\sim 10^{14}$  pN/s, weaker systems may be unfolded in the deterministic regime. In this case, a slower pulling velocity would have to be chosen, requiring considerably more computational time.

#### Investigation of the Low Force Resilience of Cohesin

1. Our simulation results in combination with calculated force propagation pathways and correlation communities suggested that the high flexibility in the region around amino acids 100–110 could be responsible for a badly formed mechanical clamp between the N- and C-terminal  $\beta$ -sheets of cohesin 1. We proposed the aforementioned mutants, A105G, P106G, T107S, and the triple mutant GGS carrying all three mutations, in an attempt to affect the folding and the formation of the mechanical clamp motif and, ideally, improve mechanical stability.

Following the same modeling and equilibration protocol followed previously, we obtained structural models for the mutants. An inspection of these structures, after 100 ns of MD, revealed how the fold can be affected by a single A105G mutation, as shown in Figure 6b. The A105G point mutation resulted in a longer mechanical clamp between the N- and C-terminal  $\beta$ -sheets. SMD simulations predicted an increase in unfolding forces both for mutant A105G and for mutant GGS, the latter being the most promising as shown in Figure 6c. It is noteworthy that the simulations revealed that the single A105G mutation already increases the force necessary to unfold cohesin 1 up to the same levels of the cohesin 3.

We want to emphasize that a strategy of carrying out only a couple of SMD simulations with low pulling velocity might give an incomplete picture of a biomolecular system under shear force. The approach adopted here, of simulating many fast pulling simulations (totaling over 350 independent SMD runs), showed that the force distribution in simulations is as widely distributed as in experiments, and therefore a small *in silico* sampling might reveal differing trends to those observed experimentally. A possibility to sample both with slow pulling and many replicas would be to employ coarse-grained methods, which are less computationally demanding. However, our simulations revealed that seemingly small mutations, like changing a methyl group to a hydrogen, can cause enormous differences in folding and therefore force resilience. With such minor changes in the biomolecule, exploratory studies to design new mutants using coarse-grained molecular dynamics simulations would be hardly reliable, as they would lack atomic detail.

In order to test the predictions from the SMD simulations, we compared the proposed mutants to the wild-type hanging cohesins experimentally (Figure 7). We found that the two promising mutants, A105G and GGS, showed a considerable increase in mechanical stability. Mutant A105G showed an increase of most probable rupture force by nearly 2.6-fold to 370 pN, relative to its wild-type cohesin 1, which unfolded at 142 pN. As predicted by the SMD simulations, this seemingly small change from an alanine to a glycine outside of the mechanical clamp motif influenced the fold of the protein enough to make it as strong as cohesin 3. The triple mutant GGS showed, again as predicted by the SMD simulations, the largest increase in unfolding force to 440 pN, making it as strong as cohesin 2, the strongest cohesin within the group of

hanging cohesins, which unfolded at 431 pN. The experimental results confirm the amino acids responsible for the low force resilience of cohesin 1 and the predicted increase in mechanostability of the proposed mutants, thus corroborating the *in silico* approach.

## CONCLUSIONS

Multienzyme molecular devices like cellulosomes rely on scaffoldins for the organization of their active constituents. Since these large protein structures can sometimes be subjected to sizable forces, their mechanical stability is a prerequisite for proper and sustained function. This holds particularly true for the scaffoldin ScaA of *A. cellulolyticus*, which anchors the microbe to a cellulose fibril through a CBM in addition to spatially organizing an ensemble of cellulolytic enzymes. Another interesting aspect is that cellulosome scaffoldins are mainly composed of highly similar cohesin domains with very different mechanical properties. Our *in vitro* and *in silico* SMFS studies not only show that the mechanical stability of all cohesins from the scaffoldin ScaA is consistent with the hypothesis proposed by Valbuena et al.<sup>13</sup> that bridging cohesins are mechanically stronger than hanging cohesins but, moreover, how minimal differences in protein sequence can lead to very different behavior under shear force.

We elucidated the surprisingly low unfolding force of cohesin 1, when compared to the other ScaA cohesins. We found that the point mutation A105G increased the mechanical stability of cohesin 1 more than 2-fold when compared to wild-type. The remarkably strong influence on the mechanical stability of cohesin 1 of a single alanine to glycine mutation, which effectively only substitutes a methyl group by a hydrogen atom, raises the question why evolutionary pressure has not favored this mutant, as it comes at virtually no additional cost for the organism. Possibly not all cohesins are supposed to display high mechanical stability, since cellulosomal organisms have already been shown to be able to regulate their gene expression patterns depending on potentially varying substrates.<sup>44,45</sup> Thus, occasionally un- and refolding cohesins would ensure that cellulosomal components can be exchanged in case of changing environmental conditions.

Both approaches, *in silico* and *in vitro*, of our combined approach started from the genetic information coding for the protein, from which the homology models for the former were derived and the samples for the latter were expressed. Given the large number of cellulosome producing microorganisms with sequenced genomes, a wide spectrum of novel combinations, for example, cohesin–dockerin pairs with similar or orthogonal affinities and tunable strengths, may be analyzed, modified, and combined. The fact that cellulosomes are extracellular organelles of microbes that live in largely diverse ambient environments, including the human gut,<sup>46,47</sup> guarantees robustness of its molecular building blocks and their interactions. This is reflected in the extremely high unfolding barriers and rupture forces of its molecular constituents and qualifies them for a large range of potential applications.

As viable candidates for source materials in a rationally designed artificial protein nanomachine, cellulosomes have demonstrated large potential in molecular engineering applications.<sup>1,6,8</sup> The development of recombinant designer cellulosomes using so-called chimeric scaffoldins allowed control over the position of each enzyme in the cellulosomal complex.<sup>7</sup> Synthetic scaffolds containing orthogonal cohesin domains have furthermore been successfully displayed on the

surface of yeast cells, allowing dockerin-tagged cellulases to bind and improve ethanol production almost by a factor of 3 compared to free enzymes.<sup>48</sup> In terms of industrial cellulose degradation, the incorporation of mechanically stronger cohesin domains and cohesin–dockerin interactions with higher affinities will make designer cellulosomes more durable and efficient. A better understanding of individual cellulosomal components can improve upcoming designs and lead to more efficient and reliable multienzyme molecular devices. For example, the new-found properties of ScaA recommend this scaffoldin and its cohesin domains to be part of a potential versatile molecular breadboard for the programmed self-assembly of molecular devices with designed properties.

From a technical point of view, we were able to measure seven constructs using a single cantilever in two separate experiments by utilizing a fast and parallelized sample preparation method, while still achieving sufficient statistics ( $N = 1420$  in 24 h and  $N = 7869$  in 72 h). Furthermore, we have shown that even in the absence of crystallized protein structures, SMD simulations, when combined with protein homology modeling, are a powerful tool to investigate the intricate mechanisms governing protein mechanics. Particularly force propagation and community analyses have proven instrumental, not only allowing us to analyze the origins of a particular molecular property, such as the low mechanical stability of cohesin 1, but also opening new means to identify crucial regions for point mutations aiming at locally altering the mechanics of the protein of choice. In summary, our newly developed methods are enabling novel investigations of protein unfolding and rational modification of structural aspects of proteins based on common design principles across different families of proteins well beyond the cellulosome community. Our results demonstrate a strategy that can be applied in fine-tuning mutations that can change the mechanostability of protein domains and also raise further questions about the evolutionary pressures that can result in mechanically stronger or weaker proteins. Considering the vast number of cellulosomal constituents yet to be explored, the combination of techniques presented here can potentially accelerate the probing and design of scaffolding domains, starting from nothing more than their genetic code, presenting new opportunities in molecular engineering and biotechnology.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b07574.

Materials and methods (gene construction, protein purification, AFM sample preparation, one-step *in vitro* expression and protein pulldown, AFM SMFS measurements, AFM SMFS data analysis, structural model determination, and molecular dynamics simulations), Supplementary Figures S1–S16, and protein sequences (PDF)

## AUTHOR INFORMATION

Corresponding Author  
\*michael.nash@unibas.ch

ORCID

Hermann E. Gaub: 0000-0002-4220-6088

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Support for this work was provided by the EU seventh Framework Programme NMP4-SL-2013-604530 (CellulosomePlus), the Nanosystems Initiative Munich, and the ERC Advanced Grant CelluFuel. M.A.N. acknowledges support from Society in Science – The Branco Weiss Fellowship from ETH Zurich, from an ERC Starting Grant (no. 715207), and from the Human Frontier Science Program (grant no. RGY0080/2015). This work was supported by National Institutes of Health (NIH) grant 9P41GM104601, “Center for Macromolecular Modeling and Bioinformatics”. R.C.B. is partially supported by the National Science Foundation (NSF) grant MCB-1616590, “Molecular Modeling of Bioenergetic Systems”, as well as the Energy Biosciences Institute (Regents of the University of California Berkeley, EBI 231 UCB BP 2014004J01). Molecular dynamics simulations used for structure prediction and equilibration made use of Blue Waters supercomputer as part of the Petascale Computational Resource (PRAC) grant “The Computational Microscope”, which is supported by the National Science Foundation (award number ACI-1440026). Blue Waters sustained-petascale computing project is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Steered molecular dynamics simulation replicas made use of the Argonne Leadership Computing Facility (ALCF)/Mira supercomputer as part of the DoE ALCC program. This research used resources of the ALCF, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. T.V. thanks Lukas F. Milles for providing the force curve analysis software, as well as E. Durner, M. A. Jobst, and W. Vanderlinden for helpful discussions and Thomas Nicolaus and Angelika Kardinal for laboratory assistance. We thank Edward A. Bayer and Marcelo Melo for helpful discussions.

## REFERENCES

- Bayer, E. A.; Morag, E.; Lamed, R. *Trends Biotechnol.* **1994**, *12* (9), 379.
- Bayer, E. a.; Lamed, R.; White, B. a.; Flint, H. J. *Chem. Rec.* **2008**, *8* (6), 364.
- Carvalho, A. L.; Dias, F. M. V.; Prates, J. A. M.; Nagy, T.; Gilbert, H. J.; Davies, G. J.; Ferreira, L. M. A.; Romão, M. J.; Fontes, C. M. G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (24), 13809.
- Ding, S.-Y.; Bayer, E. A.; Steiner, D.; Shoham, Y.; Lamed, R. *J. Bacteriol.* **1999**, *181* (21), 6720.
- Ding, S.-Y.; Xu, Q.; Crowley, M.; Zeng, Y.; Nimlos, M.; Lamed, R.; Bayer, E. A.; Himmel, M. E. *Curr. Opin. Biotechnol.* **2008**, *19* (3), 218.
- Xu, Q.; Ding, S.-Y.; Brunecky, R.; Bomble, Y. J.; Himmel, M. E.; Baker, J. O. *Biotechnol. Biofuels* **2013**, *6* (1), 126.
- Fierobe, H.-P.; Mingardon, F.; Mechaly, A.; Bélaich, A.; Rincon, M. T.; Pagès, S.; Lamed, R.; Tardif, C.; Bélaich, J.-P.; Bayer, E. a. *J. Biol. Chem.* **2005**, *280* (16), 16325.
- Gefen, G.; Anbar, M.; Morag, E.; Lamed, R.; Bayer, E. a. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (26), 10298.
- Khan, A. W. *Microbiology* **1980**, *121* (2), 499.
- Pandya, P. R.; Singh, K. M.; Parnerkar, S.; Tripathi, A. K.; Mehta, H. H.; Rank, D. N.; Kothari, R. K.; Joshi, C. G. *J. Appl. Genet.* **2010**, *51* (3), 395.
- Dassa, B.; Borovok, I.; Lamed, R.; Henrissat, B.; Coutinho, P.; Hemme, C. L.; Huang, Y.; Zhou, J.; Bayer, E. a. *BMC Genomics* **2012**, *13* (1), 210.
- Hamberg, Y.; Ruimy-Israeli, V.; Dassa, B.; Barak, Y.; Lamed, R.; Cameron, K.; Fontes, C. M. G. a.; Bayer, E. a.; Fried, D. B. *PeerJ* **2014**, *2*, No. e636.
- Valbuena, A.; Oroz, J.; Hervás, R.; Vera, A. M.; Rodríguez, D.; Menéndez, M.; Sulkowska, J. L.; Cieplak, M.; Carrión-Vázquez, M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13791.
- Mechaly, A.; Fierobe, H.-P.; Bélaich, A.; Bélaich, J.-P.; Lamed, R.; Shoham, Y.; Bayer, E. A. *J. Biol. Chem.* **2001**, *276* (13), 9883.
- Stahl, S. W.; Nash, M. a.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. a.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (50), 20431.
- Florin, E. L.; Moy, V. T.; Gaub, H. E. *Science* **1994**, *264* (5157), 415.
- Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276* (5315), 1109.
- Rico, F.; Gonzalez, L.; Casuso, I.; Puig-Vidal, M.; Scheuring, S. *Science* **2013**, *342* (6159), 741.
- King, G. M.; Carter, A. R.; Churnside, A. B.; Eberle, L. S.; Perkins, T. T. *Nano Lett.* **2009**, *9* (4), 1451.
- Müller, D. J.; Engel, A. *Nat. Protoc.* **2007**, *2* (9), 2191.
- Fernandez, J. M.; Li, H. *Science* **2004**, *303* (5664), 1674.
- Walder, R.; LeBlanc, M.-A.; Van Patten, W. J.; Edwards, D. T.; Greenberg, J. A.; Adhikari, A.; Okoniewski, S. R.; Sullan, R. M. A.; Rabuka, D.; Sousa, M. C.; Perkins, T. T. *J. Am. Chem. Soc.* **2017**, *139* (29), 9867.
- Otten, M.; Ott, W.; Jobst, M. a.; Milles, L. F.; Verdorfer, T.; Pippig, D. a.; Nash, M. a.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127.
- Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. *Nano Lett.* **2015**, *15* (11), 7370.
- Seppälä, J.; Bernardi, R. C.; Haataja, T. J. K.; Hellman, M.; Pentikäinen, O. T.; Schulten, K.; Permi, P.; Yläanne, J.; Pentikäinen, U. *Sci. Rep.* **2017**, *7* (1), 4218.
- Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. a.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. a.; Schulten, K.; Gaub, H. E.; Nash, M. a. *Nat. Commun.* **2014**, *5*, 5635.
- Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 81.
- Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. *J. Phys. Chem. B* **2017**, *121*, 3620.
- Brand, U.; Gao, S.; Engl, W.; Sulzbach, T.; Stahl, S. W.; Milles, L. F.; Nesterov, V.; Li, Z. *Meas. Sci. Technol.* **2017**, *28* (3), 034010.
- Yin, J.; Lin, A. J.; Golan, D. E.; Walsh, C. T. *Nat. Protoc.* **2006**, *1* (1), 280.
- Izrailev, S.; Stepaniants, S.; Balsara, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72* (4), 1568.
- Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72* (4), 1541.
- Goh, B. C.; Hadden, J. A.; Bernardi, R. C.; Singharoy, A.; McGreevy, R.; Rudack, T.; Cassidy, C. K.; Schulten, K. *Annu. Rev. Biophys.* **2016**, *45*, 253.
- Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Curr. Protoc. Protein Sci.* **2007**, 2.9.1.
- Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215* (3), 403.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. In *Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; International Tables for Crystallography, Vol. F; Springer: Netherlands, 2006; pp 675–684.
- Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781.
- Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. *Sci. Rep.* **2016**, *6*, 26536.
- Izrailev, S.; Stepaniants, S.; Israilewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. In *Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer: Berlin, Heidelberg, 1999; pp 39–65.
- Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96* (10), 108101.

- (41) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6620.
- (42) Lange, O. F.; Grubmüller, H. *Proteins: Struct., Funct., Genet.* **2006**, *62* (4), 1053.
- (43) Zhang, Y. *Curr. Opin. Struct. Biol.* **2008**, *18* (3), 342.
- (44) Riederer, A.; Takasuka, T. E.; Makino, S.-I.; Stevenson, D. M.; Bukhman, Y. V.; Elsen, N. L.; Fox, B. G. *Appl. Environ. Microbiol.* **2011**, *77* (4), 1243.
- (45) Han, S. O.; Yukawa, H.; Inui, M.; Doi, R. H. *J. Bacteriol.* **2003**, *185* (20), 6067.
- (46) Chassard, C.; Delmas, E.; Robert, C.; Lawson, P. A.; Bernalier-Donadille, A. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 138.
- (47) Morais, S.; Ben David, Y.; Bensoussan, L.; Duncan, S. H.; Koropatkin, N. M.; Martens, E. C.; Flint, H. J.; Bayer, E. A. *Environ. Microbiol.* **2016**, *18* (2), 542.
- (48) Tsai, S.-L.; Goyal, G.; Chen, W. *Appl. Environ. Microbiol.* **2010**, *76* (22), 7514.

## Supporting Information for

### Combining *in Vitro* and *in Silico* Single Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics

Tobias Verdorfer<sup>1</sup>, Rafael C. Bernardi<sup>2</sup>, Aylin Meinhold<sup>1</sup>, Wolfgang Ott<sup>1</sup>, Zaida Luthey-Schulten<sup>2,3</sup>, Michael A. Nash\*<sup>4,5</sup> and Hermann E. Gaub<sup>1</sup>

<sup>1</sup> Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany.

<sup>2</sup> Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States.

<sup>3</sup> Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States.

<sup>4</sup> Department of Chemistry, University of Basel, 4056 Basel, Switzerland.

<sup>5</sup> Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland.

#### Author Contributions:

TV: experiment design, sample preparation, measurements, data analysis, writing of manuscript; RCB: SMD simulations and analysis, writing of manuscript; AM: sample preparation, measurements; WO: protein design and expression; ZLS: simulation analysis, writing of manuscript; MAN: experiment design, writing of manuscript; HEG: experiment design, writing of manuscript;

\*Correspondence and requests for materials should be addressed to M.A.N. (email: [michael.nash@unibas.ch](mailto:michael.nash@unibas.ch)).

**KEYWORDS** Single molecule force spectroscopy, cohesin-dockerin, cellulosome, steered molecular dynamics, network analysis

## Materials and Methods

All reagents were at least of analytical purity grade and all buffers were filtered using a 0.2  $\mu\text{m}$  polyethersulfone membrane filter (Nalgene, Rochester, NY, USA) prior to use. All incubation steps were done at room temperature, if not otherwise stated.

### Gene construction, protein expression and purification

All genes were codon optimized for *E. coli* and synthesized (Invitrogen GeneArt Gene Synthesis - Thermo Fisher Scientific Messtechnik GmbH, Regensburg, Germany). All constructs were cloned into pET28a vectors using the Gibson assembly strategy<sup>1</sup> (New England Biolabs, MA, USA). All protein sequences can be found in the Supplementary Information.

ScaA cohesin mutant plasmid DNA was constructed using individually designed primers (Eurofins Genomics GmbH, Ebersberg, Germany) and the Phusion High-Fidelity PCR Kit (Thermo Fisher Scientific Messtechnik GmbH, Regensburg, Germany). The resulting double stranded linear DNA was ligated, phosphorylated and the template DNA was digested, in parallel, using a homemade reaction mix (1 $\mu\text{l}$  CutSMART buffer, 1 $\mu\text{l}$  ATP, 1 $\mu\text{l}$  T4 Polynucleotide Kinase, New England Biolabs, MA, USA, 1 $\mu\text{l}$  *DpnI*, 1 $\mu\text{l}$  T4 DNA ligase and 0.5 $\mu\text{l}$  PEG-6k, Thermo Fisher Scientific GmbH, Regensburg, Germany, combined with 4.5 $\mu\text{l}$  unpurified PCR product) incubated at 37°C for 15min, 22°C for 45min and finally at 80°C for 5min. All plasmids used in *in vitro* protein expression were amplified in DH5-alpha cells, purified using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany), eluted with ultrapure water and stored at -20° C. All sequences were finally checked by DNA sequencing (Eurofins Genomics GmbH, Ebersberg, Germany).

Coh3-ddFLN<sub>4</sub>-HIS-ybbR protein was expressed in *E. coli* NiCo21(DE3) cells (New England Biolabs, MA, USA). Precultures of 5 mL in LB medium, grown overnight at 37°C, were inoculated in ZYM-5052 auto-induction media containing kanamycin and grown for 6 h at 37°C followed by 24 h at 25°C<sup>2</sup>. Bacteria were spun down, and stored at -80°C. The pellet was resuspended and cells were lysed through sonication followed by centrifugation at 18000 g for 1 h at 4°C. The supernatant was applied to a Ni-NTA column (GE Healthcare, MA, USA) for HIS-Tag purification and washed extensively using HIS wash buffer (25mM TRIS, 500mM NaCl, 0.25% Tween-20, 10 % (v/v) Glycerol, 20mM imidazole, pH 8.5 @ 4°C), followed by a elution using HIS elution buffer (HIS wash buffer with 200mM imidazole instead of 20mM). Fractions containing protein were concentrated over regenerated cellulose filters (Amicon, Merck KGaA, Darmstadt, Germany), exchanged into measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl<sub>2</sub>, pH 7.2) using desalting columns (Zeba, Thermo Scientific, MA, USA), and frozen with 25 % (v/v) glycerol in liquid nitrogen to be stored at -80°C until used in experiments.

### AFM Sample preparation

The sample preparation in these experiments follows in principle previously published protocols.<sup>3-5</sup> In brief, both the AFM cantilevers (Biolever Mini, Olympus, Tokyo, Japan) and the microscope slides (76mmx26mm, Carl Roth GmbH, Karlsruhe, Germany) were cleaned and silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES, abcr GmbH, Karlsruhe, Germany) by baking at 80°C for 1h. A multiwell mask (CultureWell Gasket, Grace Bio-Labs, Bend, USA) was cleaned by sonication in a 1:1 mixture of isopropyl alcohol (IPA) and ultrapure water and then dried in a stream of nitrogen. The mask was attached to the glass slide to allow compartmentalization of the surface. The cantilevers were incubated with heterobifunctional NHS-PEG-Maleimide (5 kDa; Rapp Polymere, Tübingen, Germany) in 100mM HEPES buffer pH 7.4 for 45 min. The surfaces in the wells however were incubated with a 1:100

mixture of NHS-PEG-Maleimide and NHS-PEG-CH<sub>3</sub> (both 5 kDa; Rapp Polymere, Tübingen, Germany) in 100mM HEPES buffer pH 7.4, which, as experience has shown, will later result in the right surface density of immobilized protein for SMFS measurements in these experiments. After rinsing with ultrapure water, both the cantilevers and the surfaces were incubated with 1 mM Coenzyme A (CoA) in a 1 mM sodium phosphate pH 7.2, 50 mM NaCl, 10 mM EDTA buffer for at least 1 h. After a final ultrapure water rinse the cantilevers were incubated with 40µM Coh3-ddFLN<sub>4</sub>-HIS-ybbR and 5 µM phosphopantetheinyl transferase (Sfp) for 2 h with magnesium chloride supplemented measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl<sub>2</sub>, 20mM MgCl<sub>2</sub> pH 7.2). The glass slide with the multiwell mask still attached was stored under Argon for later use. The cantilevers were rinsed extensively with measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl<sub>2</sub>, pH 7.2) and finally stored in it until use in measurement.

## One-step *in vitro* expression and protein pulldown

PURExpress® IVTT-kit was thawed on ice and supplemented with 5 µM Sfp, 0.8 U/µl RNase inhibitor (NEB #M0314), 10 ng/µl Plasmid-DNA, 0.05% v/v Triton X-100 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) and ultrapure water resulting in total volumes of 10µl for each reaction mix. There was no need to supply this reaction mix with additional MgCl<sub>2</sub> for the Sfp coupling reaction, since the PURExpress® IVTT-kit already contains 13mM MgCl<sub>2</sub><sup>6</sup>. The reaction mixes were transferred to the wells onto the CoA functionalized glass slide and incubated at 37°C for 3h. During this time the cell free expression kit is constantly producing proteins, while at the same time the Sfp couples expressed protein to the surface via the ybbR tags. Both cantilevers and surfaces were rinsed extensively with measurement buffer (TBS- Ca: 25 mM Tris, 72 mM NaCl, 1mM CaCl<sub>2</sub>, pH 7.2) before measurement and finally the multiwell mask was removed from the surface and stored in a 1:1 mixture of IPA and ultrapure water for further use.

## AFM SMFS measurements

A custom build AFM connected to a MFP3D controller (Asylum Research, Santa Barbara, CA, USA) was used for all measurements. Acquisition- and instrument control software was written in Igor Pro 6 (Wavemetrics, OR, USA). The cantilever was aligned to each measurement spot by moving the sample using a 25mmx25mm piezomotor stage (PI, Karlsruhe, Germany) using a camera mounted below the sample. These positions were saved in the software for later use. The cantilever was brought in close proximity to the surface and constant speed measurements with retraction speeds of 1600 nm/s were started. The glass surface was moved horizontally by 100 nm in a snail-like-pattern within each protein spot. After 2000 approach- and retract-cycles the AFM-head was automatically lifted by a linear piezo actuator (Newport, CA, USA) and the surface was moved horizontally by typically ~300 µm to expose the cantilever to the next protein spot. Cantilevers were calibrated using the equipartition theorem method<sup>7</sup>.

## AFM SMFS Data Analysis

Data analysis was carried out following previous work<sup>8</sup>. In short, data were transformed into physical units and corrected for cantilever bending, laser spot- and baseline-drift. Force peaks and rupture events were detected and transformed to contour length space. The Worm Like Chain model (WLC)<sup>9</sup> was used to fit relevant peaks. All curves showing a ddFLN<sub>4</sub> and cohesin contour length increment (ddFLN<sub>4</sub>: 34nm<sup>8,10</sup>, cohesin: 45nm) were used to assemble unfolding force histograms, which were then fitted following the Bell-Evans model<sup>11,12</sup>, which is commonly used to estimate the distance to the transition state  $\Delta x$  and the natural off-rate  $k_0$  of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments.

Bell-Evans probability density function at given loading rate  $r$ :

$$p(F) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 k_B T}{r \Delta x} \left(e^{\frac{\Delta x}{k_B T} F} - 1\right)\right]$$

The Bell-Evans model predicts a linear dependence between the most probable rupture force  $\langle F \rangle$  and the logarithm of the force loading rate  $r$ :

$$\langle F(r) \rangle = \left(\frac{k_B T}{\Delta x}\right) \ln\left(\frac{r \Delta x}{k_0 k_B T}\right)$$

The Dudko-Hummer-Szabo model<sup>13</sup> describes a non-linear dependence for the most probable rupture force on loading rate:

$$\langle F(r) \rangle = \frac{\Delta G}{\nu \Delta x} \left\{ 1 - \left[ \frac{k_B T}{\Delta G} \ln \left( \frac{k_0 k_B T}{r \Delta x} \exp \left( \frac{\Delta G}{k_B T} + \gamma \right) \right) \right]^\nu \right\}$$

where  $\Delta G$  is the free energy of activation and  $\gamma \approx 0.577\dots$  is the Euler-Mascheroni constant. The model parameter  $\nu$  defines the single-well free-energy surface model used ( $\nu = 2/3$  for linear-cubic and  $\nu = 1/2$  for cusp free-energy).

## Structural Model Determination

The amino acid sequence of all seven cohesins under investigation were obtained from the GenBank (GenBank: AAF06064.1) proteomic server<sup>14,15</sup>. The template search was performed employing the similarity search algorithm in the protein Blast server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)<sup>16</sup> using the Protein Data Bank<sup>17</sup> (<http://www.pdb.org>) as database and the default options. Using VMD's<sup>18</sup> multiseq<sup>19</sup> analysis tool, sequences were aligned to templates employing ClustalW algorithm<sup>20</sup>. The construction of cohesin models were performed using MODELLER 9.17 software<sup>21</sup> that employs spatial restriction techniques based on the 3D-template structure. The best model was selected by analyzing the stereochemical quality check using PROCHECK<sup>22</sup> and overall quality by ERRAT server.<sup>23</sup> All structures were subjected to 100 ns of equilibrium MD, as described below, to ensure conformational stability. All structures shown in this manuscript are from post-equilibration simulations.

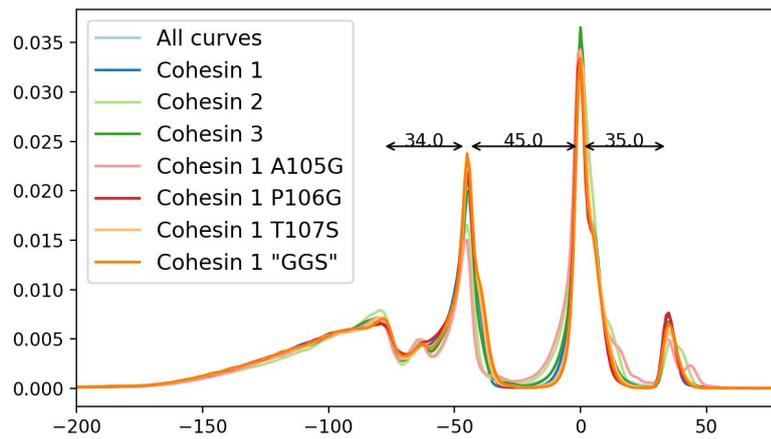
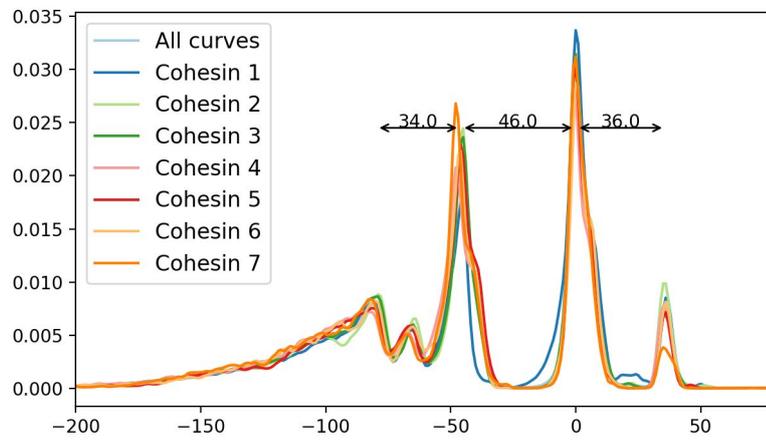
## Molecular dynamics simulations

Employing advanced run options of QwikMD,<sup>24</sup> structural models were solvated and the net charge of the proteins were neutralized using a 75 mM salt concentration of sodium chloride, which were randomly arranged in the solvent. The overall number of atoms included in MD simulations varied from 50,000 in the equilibrium simulations to near 300,000 in the pulling simulations. The MD simulations in the present study were performed employing the NAMD molecular dynamics package.<sup>25</sup> The CHARMM36 force field<sup>26,27</sup> along with the TIP3 water model<sup>28</sup> was used to describe all systems. The simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME)<sup>29</sup> method. The equations of motion were integrated using the r-RESPA multiple time step scheme<sup>25</sup> to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. Before the MD simulations all the systems were submitted to an energy minimization protocol for 1,000 steps. MD simulations with position restraints in the protein backbone atoms were performed for 10 ns and served to pre-equilibrate systems before the 100 ns equilibrium MD runs, which served to evaluate structural model stability. During the 10 ns pre-equilibration the initial temperature was set to zero and was constantly increased by 1 K every 1,000 MD steps until the desired temperature (300 K) was reached.

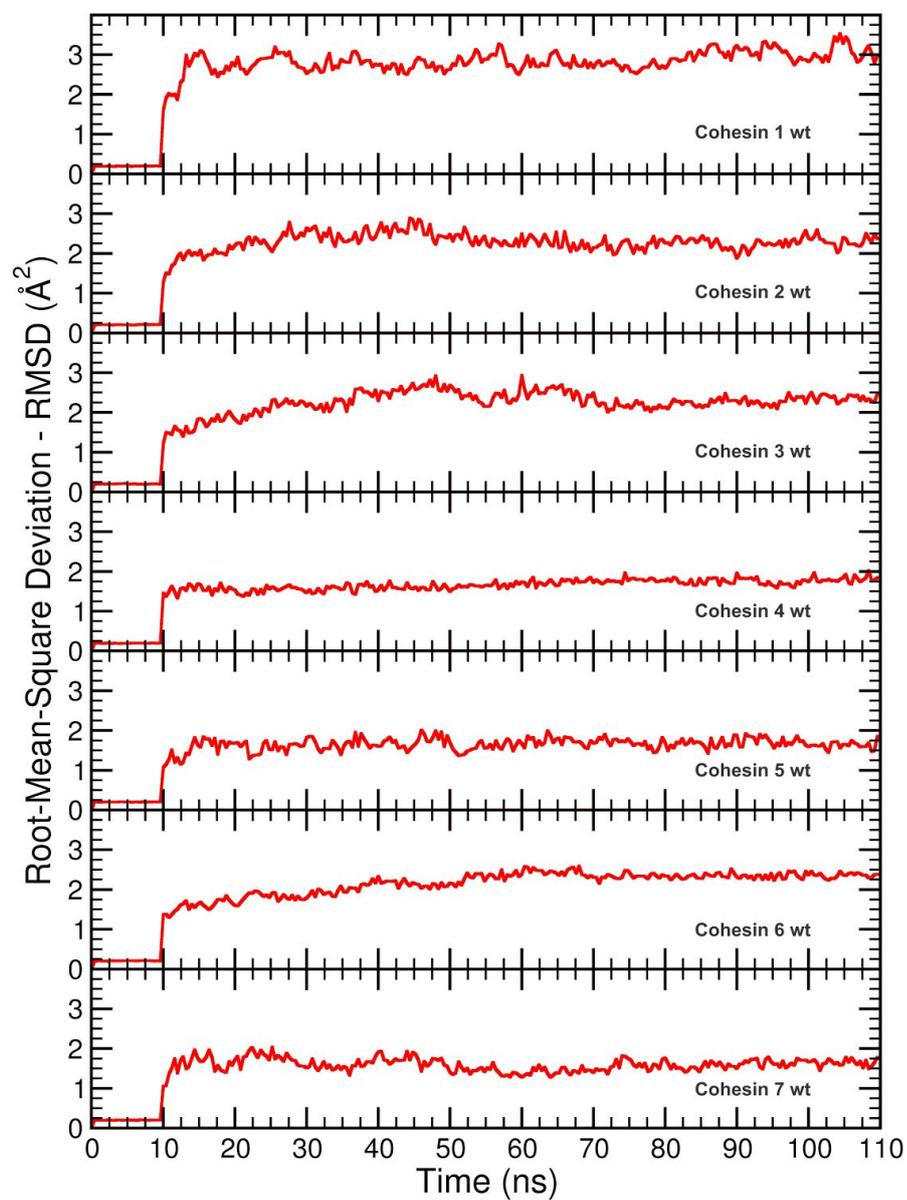
With structures properly equilibrated and checked, solvent boxes were enlarged in the Z coordinate to allow space for protein unfolding during SMD simulations. The new solvent boxes were equilibrated for 10 ns keeping the protein atoms restrained in space. SMD simulations<sup>11</sup> were performed using a constant velocity stretching (SMD-CV protocol), employing four different pulling speeds: 250, 25, 2.5 and 0.5 Å/ns.

Simulation replicas (at least 25 per system), used in all the plots in this manuscript, were performed with constant pulling speed of 2.5 Å/ns. Values for force over the pulling spring were saved every 50 steps. The spring constant of the pulling spring was set to 5.0 kcal/mol/Å<sup>2</sup>, while the holding spring had a constant of 10 kcal/mol/Å<sup>2</sup>. In all simulations, totaling over 350 SMD simulations, SMD was employed by harmonically restraining the position of N-terminal amino acid residue of the cohesin domain, and moving a second restraint point, at the C-terminal of the cohesin domain, with constant velocity in the +z direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic spring is then monitored during the time of the molecular dynamics simulation. The pulling point was moved with constant velocity along the z-axis and due to the single anchoring point and the single pulling point the system is quickly aligned along the z-axis. Owing to the flexibility of the linkers between the cohesins and fingerprint domains, this approach reproduces the experimental set-up. All analyses of MD trajectories were carried out employing VMD<sup>18</sup> and its plug-ins. Secondary structures were assigned using the Timeline plug-in, which employs STRIDE criteria.<sup>30</sup>

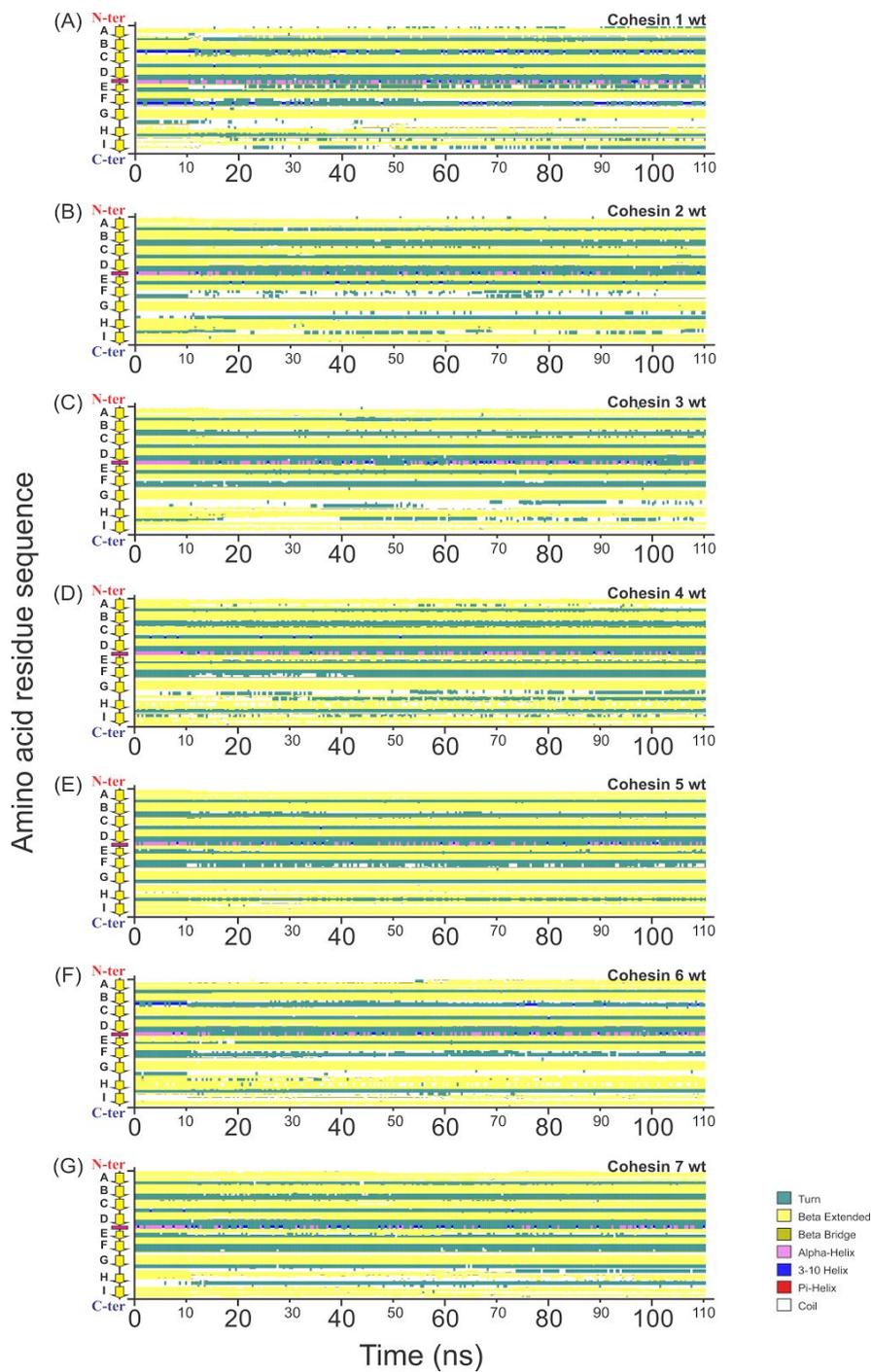
The Network View plugin<sup>31</sup> on VMD was employed to perform dynamical network analysis. A network was defined as a set of nodes, all  $\alpha$ -carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 non-consecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (non-hydrogen) from the 2 monomers are within 4.5Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.<sup>31</sup> nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths, which can be understood as allosteric signaling pathways or force propagation pathways<sup>32</sup>. Suboptimal paths are defined as paths that are slightly longer than the optimal path, with a given suboptimal path visiting a node not more than once. These multiple communication paths are nearly equal in length, and not all residues along these paths need be considered important for allostery. Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective pathway for allosteric communication. The thickness of the edges connecting the nodes reveals the least and most used paths. Allostery can be understood in terms of pathways of residues that efficiently transmit energy, here in the form of mechanical stress, between different binding sites<sup>33</sup>. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as  $w_{ij} = -\log(|C_{ij}|)$ , where  $C_{ij}$  is the correlation matrix calculated with Carma<sup>34</sup>. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was  $-\log(0.5) = 0.69$ . To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently. As previously investigated by our group<sup>32</sup>, Pearson correlation is ideal for force propagation calculation. However, due to its nature, communities analysis would benefit from an information-theory-based method, so here we employed generalized correlation<sup>35</sup> to the community analysis. Tightly correlated groups of atoms are clustered into communities, indicating functional domains of biomolecules and important interfaces between multi-molecule complexes.



**Figure S1:** Frequency of observed relative contour lengths increments determined by transforming multiple force traces into contour length space via the worm-like chain model and aligning them. The individual increments (f.l.t.r.) correspond to the unfolding of the ddFLN4 fingerprint domain, the ScaA cohesins and the occasional unfolding of the X-module of the Coh3.XDoc3 complex<sup>36</sup>.

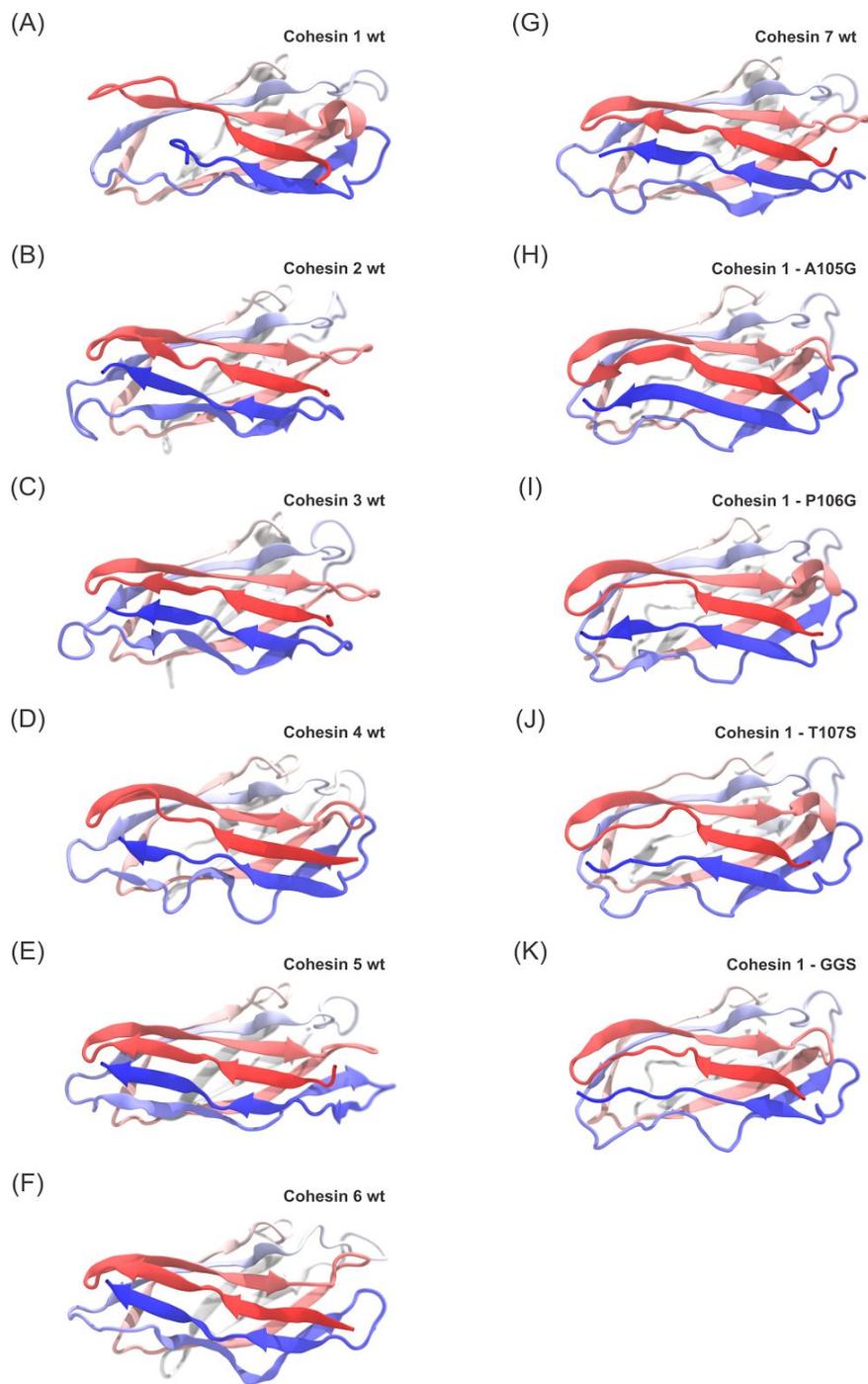


**Figure S2: Root-Mean-Square Deviation (RMSD) for equilibrium simulations.** All constructs were simulated with position restraints of the backbone atoms during 10 ns and free of restraints during 100 ns. All plots show stable structures after approximately 30ns. It is noteworthy that hanging cohesins have a higher RMSD value, particularly cohesin 1.

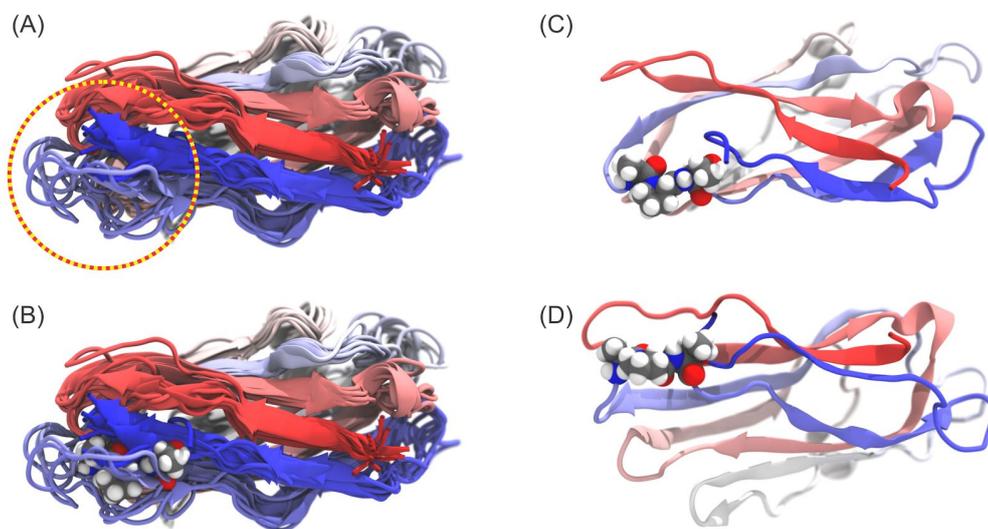


**Figure S3: Secondary Structure evolution during equilibration simulations.** Secondary structure content was evaluated using VMD's Timeline during the equilibration simulations. All constructs were simulated with position restraints of the backbone atoms during 10 ns and free of restraints during 100 ns. All conformations show stable structures.

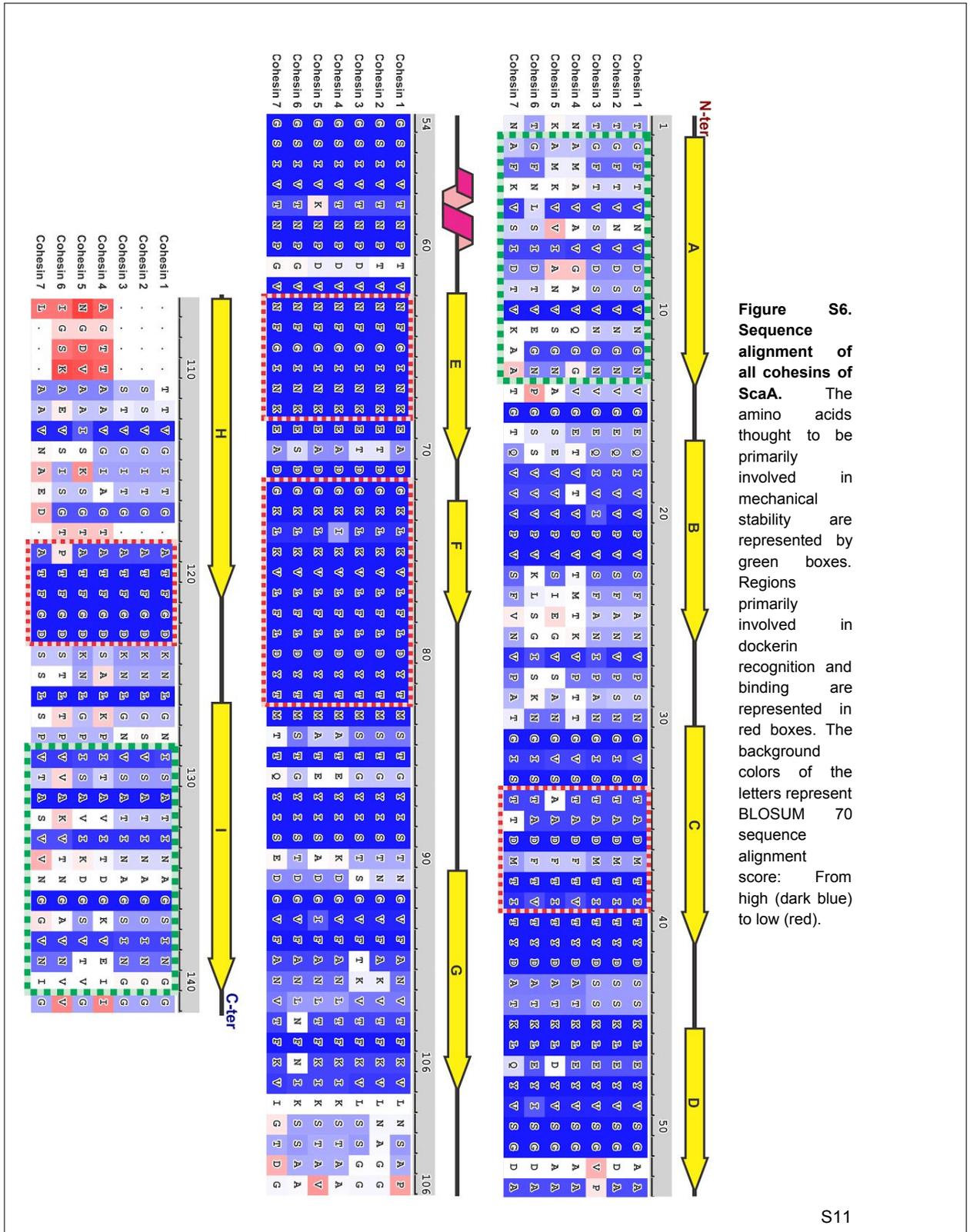
S8



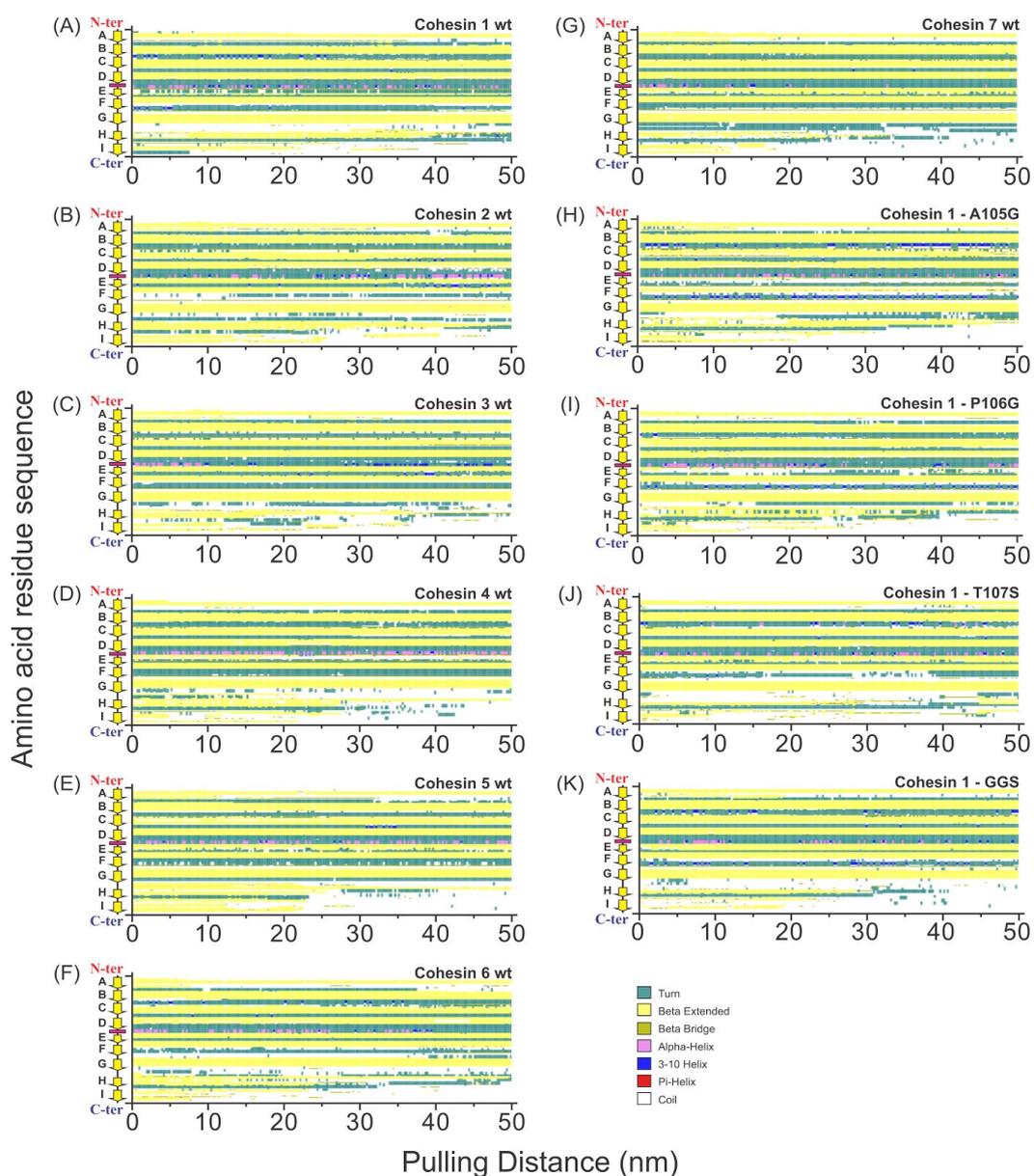
**Figure S4. Structural model of studied cohesins after 100ns of MD simulation.** All structures were obtained using Modeller 9.17 and subjected to 100 ns of molecular dynamics equilibration using QwikMD and NAMD. All images were prepared using VMD.



**Figure S5. Structural model for ScaA cohesins after 100ns of MD simulation.** (A) Using Modeller, very similar model structures were obtained for ScaA cohesins. The region highlighted in the circle was observed to be the most flexible one in cohesin 1, presenting a different behavior than the other cohesins. (B) Sequence alignment shows that, in the highlighted region of (A), 3 amino acid residues of cohesin 1 were different, compared to cohesin 2 and 3, namely ALA105, PRO106 and THR107. (C) (D) Two different viewpoints of cohesin 1 with highlighted ALA105, PRO106 and THR107.

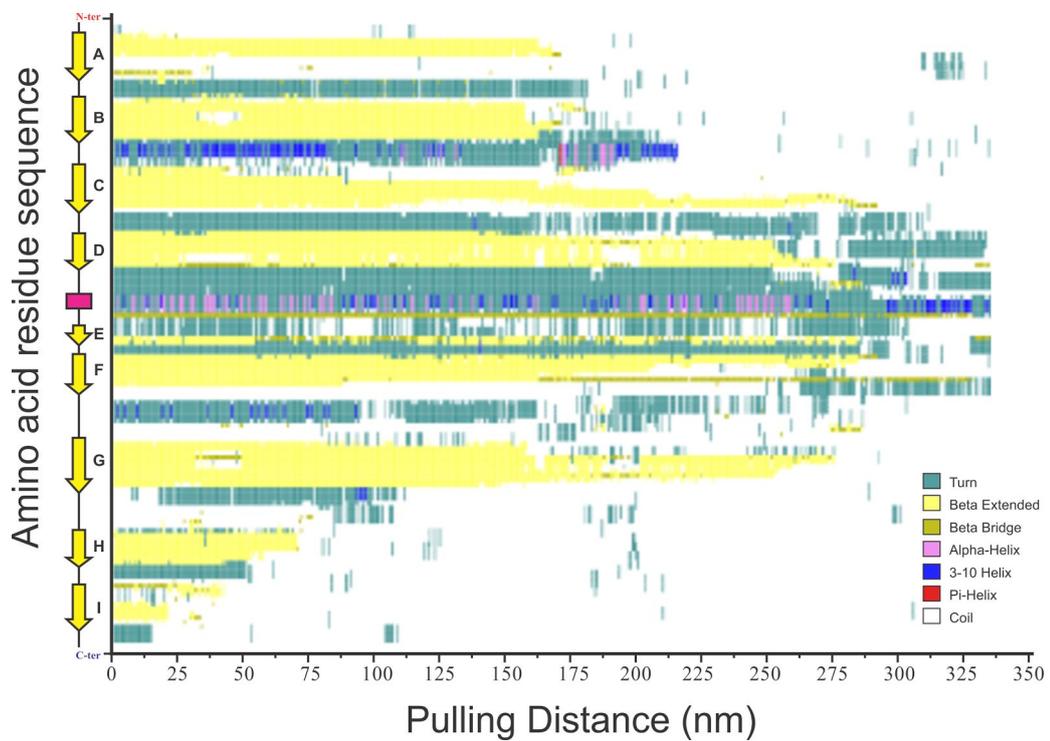


S11



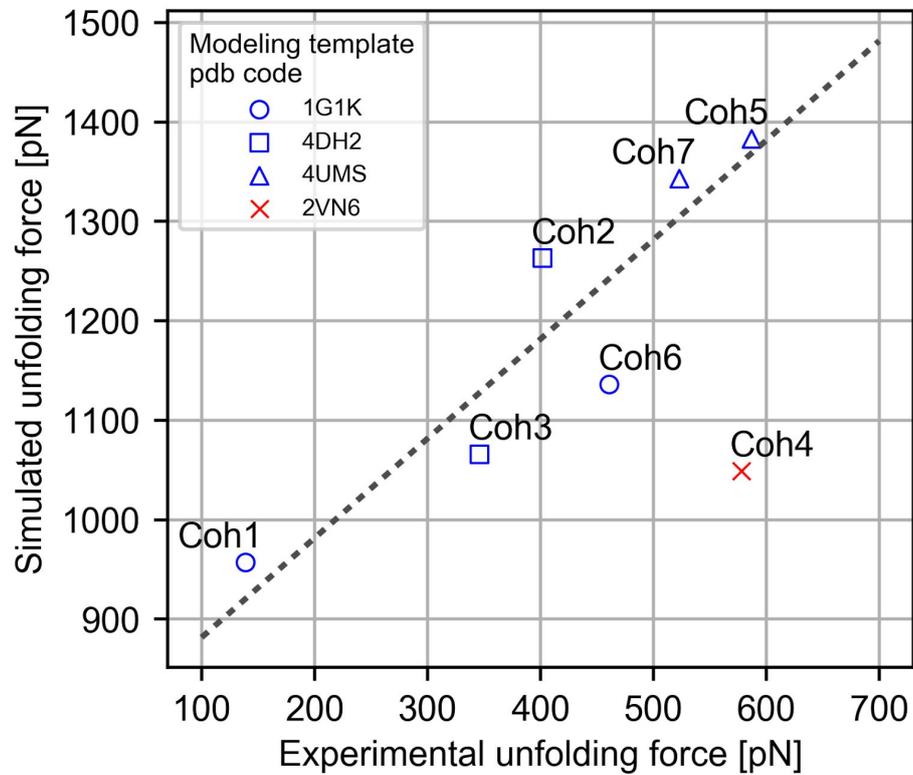
**Figure S7. Secondary Structure evolution during pulling simulations.** Secondary structure content was evaluated using VMD's Timeline during the SMD simulations. Here we show the evolution of the secondary structure during the first 50nm of pulling, which corresponds to the region where the peak force is observed. The plots present a representative simulation (one of the replicas) for each system studied. In all simulations presented the C-terminal region is the first to unfold, showing that the highest peak corresponds to losing the last beta-strand structure.

S12

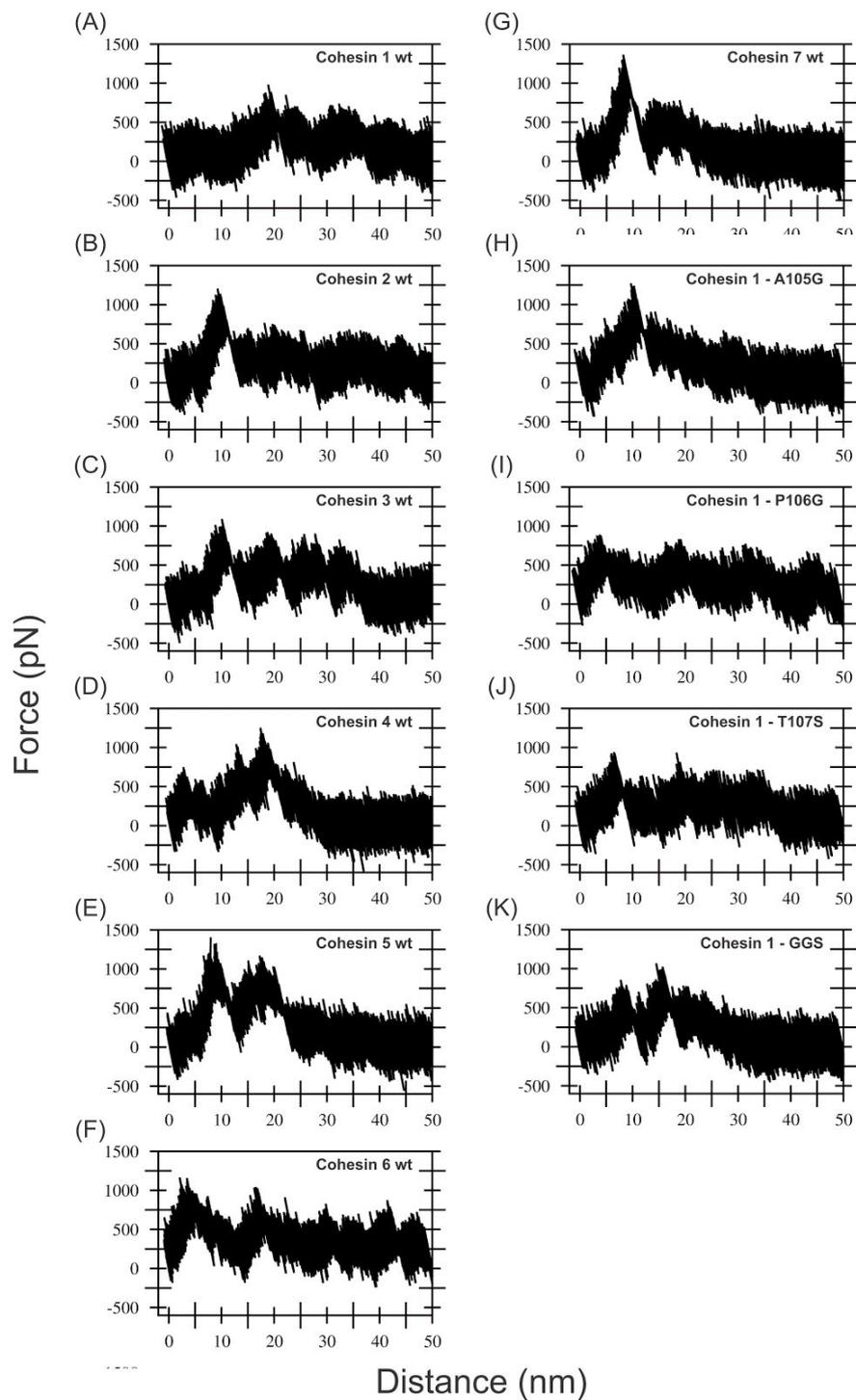


**Figure S8. Cohesin 1 secondary structure evolution during pulling simulations.** Secondary structure content was evaluated using VMD's Timeline during the SMD simulations. Here we show the evolution of the secondary structure during the whole unfolding process.

S13



**Figure S9. Experimental unfolding force vs. simulated peak unfolding force of all wild type cohesins.** A linear fit shows a clear trend between measured and simulated unfolding forces with offset fitting parameter ( $\pm$ SD)  $F_0 = 782 \pm 29$  pN. Cohesin 4 (red) was excluded from the fit since, as explained in the main text, its homology modeling template was a suboptimal and likely resulted in a non-ideal initial folded state.



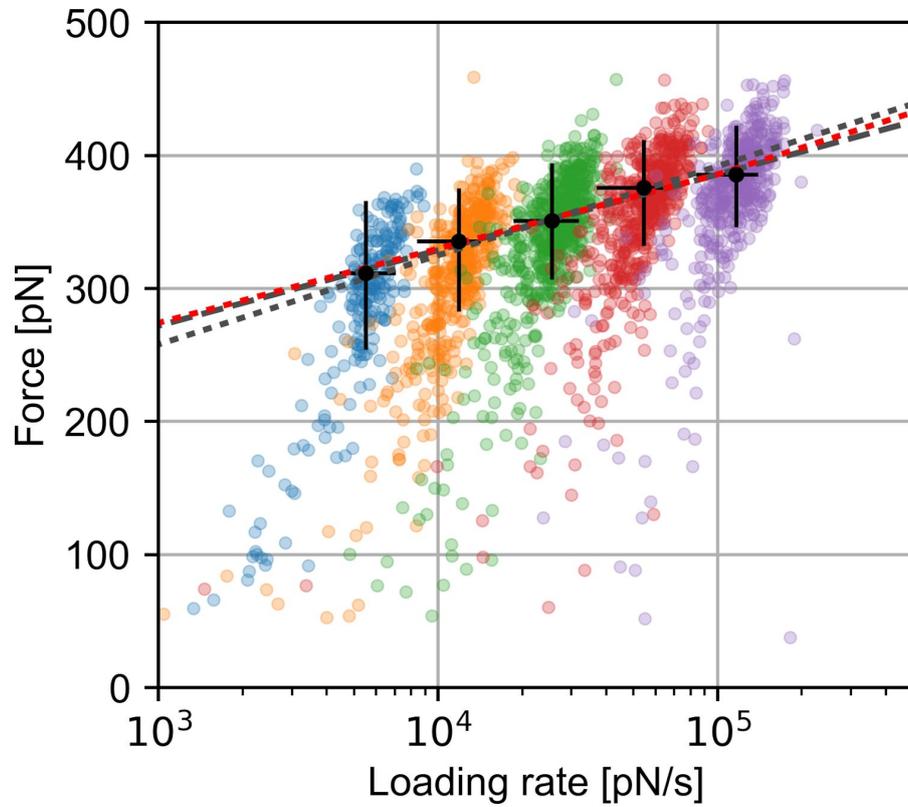
**Figure S10. Force profile during pulling simulations.** For the first 50 nm of pulling, the plots present a representative simulation (one of the replicas) for each system studied.

S15

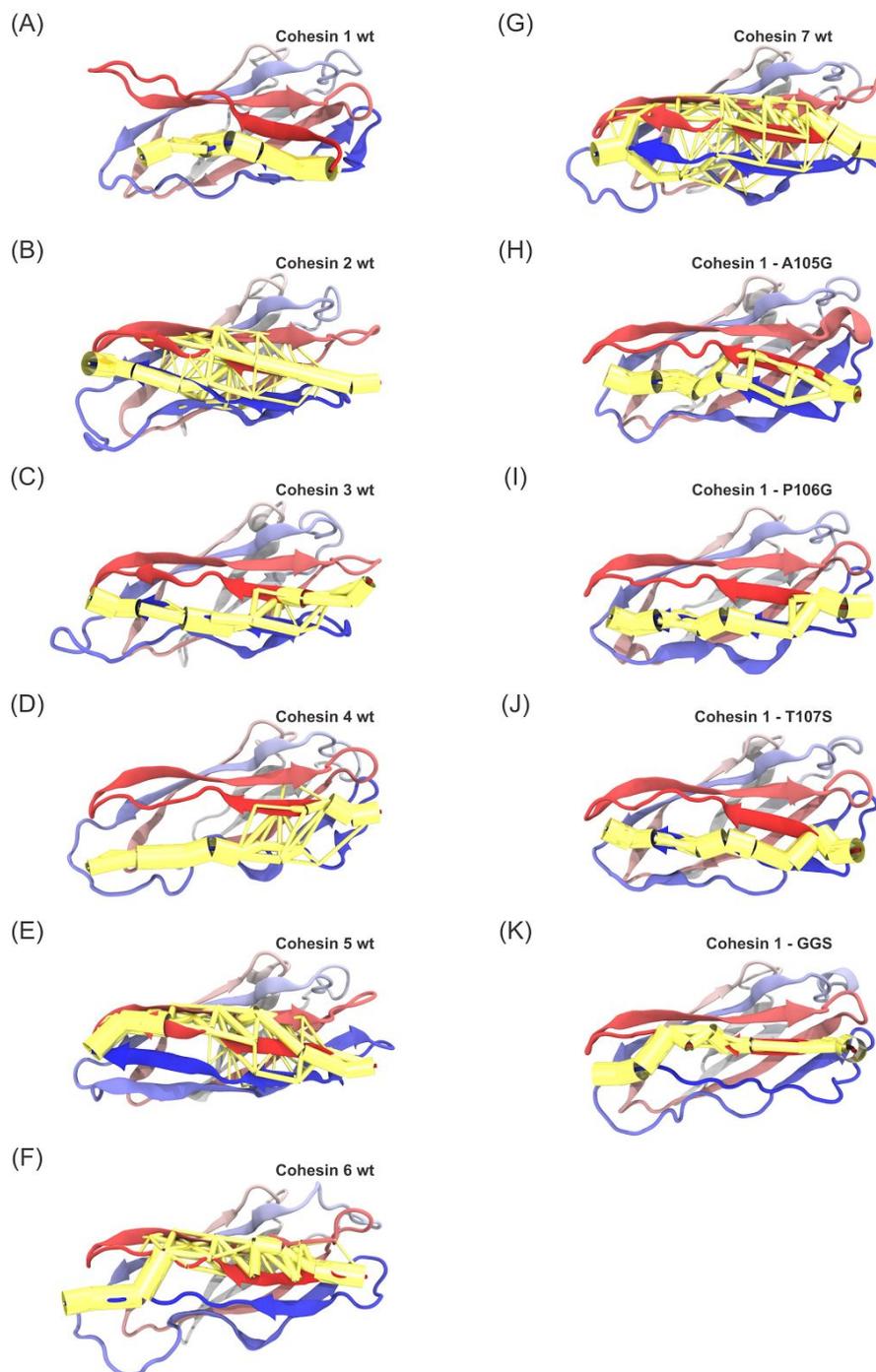
	Coh1	Coh2	Coh3	Coh4	Coh5	Coh6	Coh7
Coh1	1	**** 4.7E-11	** 0.00949	** 0.0072	**** 1.13E-10	**** 1.61E-07	**** 1.13E-10
Coh2	**** 4.7E-11	1	**** 4.14E-08	**** 1.51E-07	** 0.004	**** 1.28E-05	0.0717
Coh3	** 0.00949	**** 4.14E-08	1	0.985	**** 3.5E-09	** 0.00222	**** 1.72E-08
Coh4	** 0.0072	**** 1.51E-07	0.985	1	**** 2.04E-10	** 0.00233	**** 9.36E-09
Coh5	**** 1.13E-10	** 0.004	**** 3.5E-09	**** 2.04E-10	1	**** 8.55E-11	0.124
Coh6	**** 1.61E-07	**** 1.28E-05	** 0.00222	** 0.00233	**** 8.55E-11	1	**** 4.15E-07
Coh7	**** 1.13E-10	0.0717	**** 1.72E-08	**** 9.36E-09	0.124	**** 4.15E-07	1

\*: p < 0.05
\*\*: p < 0.01
\*\*\*: p < 0.001
\*\*\*\*: p < 0.0001

**Figure S11. Statistical significance between the simulated unfolding forces of all wild type cohesins.** P-values were calculated using the Kolmogorov-Smirnov test.

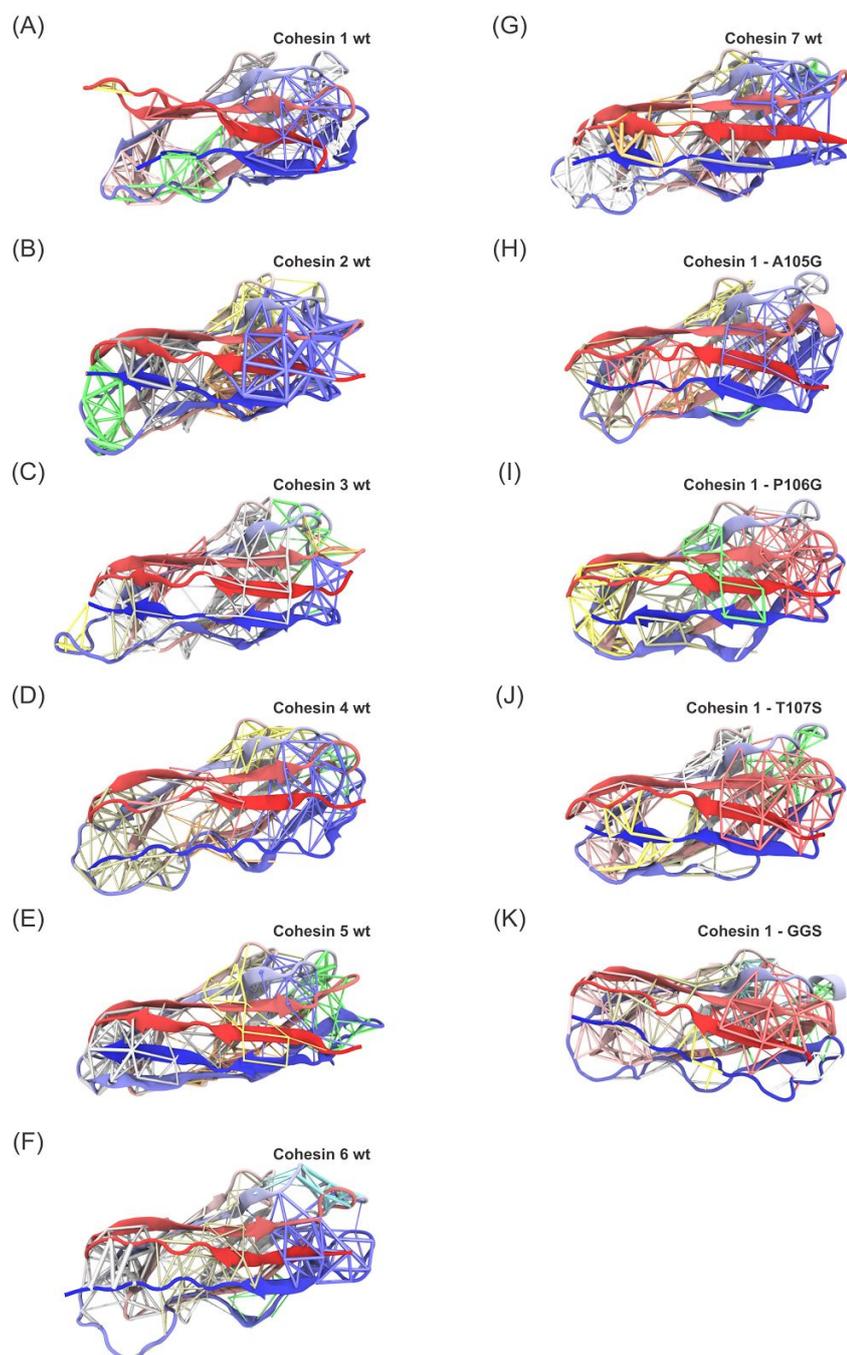


**Figure S12. Experimental dynamic force spectrum for unfolding events of cohesin 3.** Varicolored points represent rupture force/loading rate data from an experiment with 5 different pulling speeds. Black points represent the most probable rupture force/loading rate of each pulling speed obtained from kernel density estimates. Error bars represent the full width at half maximum. Gray lines represent least-squares fits of the Bell-Evans model to the experimental, and to both the experimental and the simulation data with fitting parameters ( $\pm$ SD)  $\Delta x = 0.17 \pm 0.012$  nm,  $k_0 = 6.7 \cdot 10^{-4} \pm 6.3 \cdot 10^{-4}$  s $^{-1}$  and  $\Delta x = 0.14 \pm 0.0015$  nm,  $k_0 = 4.9 \cdot 10^{-3} \pm 8.9 \cdot 10^{-4}$  s $^{-1}$ , respectively. The red dotted line represents a least-squares fit of the DHS model to both the combined experimental and the simulation data with fitting parameters ( $\pm$ SD) of  $\Delta x = 0.19 \pm 0.024$  nm,  $k_0 = 1.4 \cdot 10^{-4} \pm 2 \cdot 10^{-4}$  s $^{-1}$  and  $\Delta G = 60 \pm 13$  k $_B$ T. In this range of loading rates the Bell-Evans fit through the experimental data falls along the DHS fit.



**Figure S13. Network-based force propagation analysis calculated using Pearson correlation.** Suboptimal force paths were calculated using VMD during the first 5nm of pulling simulation. Note that for all the systems beta-strands A,B and I are the main regions involved in the force propagation.

S18



**Figure S14. Network-based community analysis calculated using generalized correlation.** Communities were calculated using VMD during the first 5nm of pulling simulation. Each color represents a different community. Colors of the communities in different systems are not related, and should not be compared as being the same community in different systems. Thickness of the network represents the  $\log$  of the normalized correlation value. Thick connections represent highly correlated regions.

S19



	Coh1	A105G	P106G	T107S	GGG
Coh1	1	0.0333 *	0.523	0.714	0.000295 ***
A105G	0.0333 *	1	0.00128 **	0.124	0.00999 **
P106G	0.523	0.00128 **	1	0.414	1.33E-05 ****
T107S	0.714	0.124	0.414	1	0.000235 ***
GGG	0.000295 ***	0.00999 **	1.33E-05 ****	0.000235 ***	1

\*: p < 0.05
\*\*: p < 0.01
\*\*\*: p < 0.001
\*\*\*\*: p < 0.0001

**Figure S16. Statistical significance between the simulated unfolding forces Cohesin 1 and its four mutants.** P-values were calculated using the Kolmogorov-Smirnov test.

## Protein Sequences

ybbR-tag - linker and additional residues - Cohesin - linker - XDoc3

MGTDSLEFIASKLALEVLFGQPLQHSHHHHPWTSAS

### Cohesin 1

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEA  
DGKLVFLDYTMSTGYISTNGVFANVTFKVLNSAPTTVGITGATFGDKNLGNISATINAGSINGG

### Cohesin 2

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGISTADMTITYDSSKLEYVSGDAGSIVTNPTVNFNGINKETD  
GKLVFLDYTMSTGYISTNGVFAKVTFKVLNAGSSVGITGATFGDKNLGVSATINAGSINGG

### Cohesin 3

TGFTVNVDSVNGNVGEQIVPVSFANIPANGISTADMTITYDSSKLEYVSGVPGSIVTNPDVNFNGINKETDG  
KLVFLDYTMSTGYISTSGVFTKVFVLSGGSTVGITGATFGDKNLGNVSATINAGSINGG

### Cohesin 4

NAMAVAVGAVQGGVGETVTPVMTKVPPTGVSTADFTVYDANKLEYVSGAAGSIVTNPDVNFNGINKEA  
DGKLVFLDYTMATEYISKDGVFANLTFKIKSTAAAGTTAAVGIAGTATFGDSALKPITAVITDGKVEII

### Cohesin 5

KAMKVIANVSGNAGSEVVVPSIEGVSANGVSAADFTITYDATKLDYVSGAAGSIVKNPDVNFNGINKEAD  
GKLVFLDYTMATEYISADGIFANLTFKIKSTAVNGDVAISKSGTATFGDKNLGPISAVIKDGSVTVG

### Cohesin 6

TGFNLSIDTVEGNPSSVVPVLSGISKNGISTADFTVYDANKLEYISGDAGSIVTNPGVNFNGINKESDG  
KLVFLDYTMSTGYISTDGVFANLNFNIKSSAAIGSKAEVSISGTPTFGDSTLTPVAVKVTNGAVNVV

### Cohesin 7

NAFKVSIDTVKAATGTQVVVPSFVNVPATGISTTDMTITYDATKLQYVSGDAGSIVTNPGVNFNGINKEAD  
GKLVFLDYTMTTQYISEDGVFANVTFKVIKTDGLAAVNAEDATFGDSSLSPTASVNVGGVNI

### Cohesin 1 A105G

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEA  
DGKLVFLDYTMSTGYISTNGVFANVTFKVLNSGPTTVGITGATFGDKNLGNISATINAGSINGG

### Cohesin 1 P106G

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEA  
DGKLVFLDYTMSTGYISTNGVFANVTFKVLNSAGTTVGITGATFGDKNLGNISATINAGSINGG

### Cohesin 1 T107S

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEA  
DGKLVFLDYTMSTGYISTNGVFANVTFKVLNSAPSTVGITGATFGDKNLGNISATINAGSINGG

### Cohesin 1 A105G P106G T107S

TGFTVNVDSVNGNVGEQIVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEA  
DGKLVFLDYTMSTGYISTNGVFANVTFKVLNSGGTVGITGATFGDKNLGNISATINAGSINGG

VVPNTVTSVAVKTQYVEIESVDGFYNTEDKFDTAQIKKAVLHTVYNEGYTGDDGVAVLREYESEPVVDITA  
ELTFGDATPANTYKAVENKFDYEIPVYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTY  
AATSTDGKDATTVALSPSTLVGGNPESVYDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFY  
KSSVDQYKDMAANEPNKLWDIVTGDAEEEE

S22

Coh3 - linker - ddFLN4 - linker and additional residues - ybbR

MGTALTRGMYDLDPKDGSAAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGYHIYWD  
ERLEVVATKTGAYAKKGAALDSSLAKAENNGVVFVASGADDFGADGVMWTVLKVLPADAKAGDVY  
PIDVAYQWDPKGLFTDNKDSAQGLKMQAYFFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGS  
VVPSTGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVHRTDGGDGFVVTIEGPAPVDPVMVDNG  
DGTYDVEFEPKEAGDYVINLTDGDNVNGFPKTVTKPAPGSELKLPSSRHHHHHHSLEVLFGQPDLSL  
EFIASKLA

## References

- (1) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. a.; Smith, H. O.; Iii, C. A. H.; America, N. *Nat. Methods* **2009**, *6* (5), 343.
- (2) Studier, F. W. *Protein Expr. Purif.* **2005**, *41* (1), 207.
- (3) Jobst, M. A.; Schoeler, C.; Malinowska, K.; Nash, M. A. *J. Vis. Exp.* **2013**, No. 82, 1.
- (4) Otten, M.; Ott, W.; Jobst, M. a.; Milles, L. F.; Verdorfer, T.; Pippig, D. a.; Nash, M. a.; Gaub, H. E. *Nat. Methods* **2014**, No. september, 1.
- (5) Zimmermann, J. L.; Nicolaus, T.; Neuert, G.; Blank, K. *Nat. Protoc.* **2010**, *5* (6), 975.
- (6) Shimizu, Y.; Kanamori, T.; Ueda, T. *Methods* **2005**, *36* (3), 299.
- (7) Hutter, J. L.; Bechhoefer, J. *Rev. Sci. Instrum.* **1993**, *64* (7), 1868.
- (8) Milles, L. F.; Bayer, E. A.; Nash, M. A.; Gaub, H. E. *J. Phys. Chem. B* **2016**, *acs.jpcc.6b09593*.
- (9) Siggia, E. D.; Bustamante, C.; Marko, J. F.; Smith, S. B. *Science* **1994**, *5*.
- (10) Schwaiger, I.; Kardinal, A.; Schleicher, M.; Noegel, A. A.; Rief, M. *Nat. Struct. Mol. Biol.* **2004**, *11* (1), 81.
- (11) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72* (4), 1568.
- (12) Evans, E.; Ritchie, K. *Biophys. J.* **1997**, *72* (4), 1541.
- (13) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96* (10), 108101.
- (14) Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Sayers, E. W. *Nucleic Acids Res.* **2017**, *45* (D1), D37.
- (15) Ding, S.-Y.; Bayer, E. A.; Steiner, D.; Shoham, Y.; Lamed, R. *J. Bacteriol.* **1999**, *181* (21), 6720.
- (16) Gish, W.; States, D. J. *Nat. Genet.* **1993**, *3* (3), 266.
- (17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. In *International Tables for Crystallography Volume F: Crystallography of biological macromolecules*; Rossmann, M. G., Arnold, E., Eds.; International Tables for Crystallography; Springer Netherlands, 2006; pp 675–684.
- (18) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33.
- (19) Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. *BMC Bioinformatics* **2006**, *7*, 382.
- (20) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. *Curr. Protoc. Bioinformatics* **2002**, Chapter 2, Unit 2.3.
- (21) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M.-Y.; Pieper, U.; Sali, A. *Curr. Protoc. Protein Sci.* **2007**, Chapter 2, Unit 2.9.
- (22) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *J. Appl. Crystallogr.* **1993**, *26* (2), 283.
- (23) MacArthur, M. W.; Laskowski, R. A.; Thornton, J. M. *Curr. Opin. Struct. Biol.* **1994**, *4* (5), 731.
- (24) Ribeiro, J. V.; Bernardi, R. C.; Rudack, T.; Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Schulten, K. *Sci. Rep.* **2016**, *6*, 26536.

S23

- (25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26* (16), 1781.
- (26) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257.
- (27) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102* (18), 3586.
- (28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926.
- (29) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089.
- (30) Frishman, D.; Argos, P. *Proteins* **1995**, *23* (4), 566.
- (31) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (16), 6620.
- (32) Schoeler, C.; Bernardi, R. C.; Malinowska, K. H.; Durner, E.; Ott, W.; Bayer, E. A.; Schulten, K.; Nash, M. A.; Gaub, H. E. *Nano Lett.* **2015**, *15* (11), 7370.
- (33) Ribeiro, A. A. S. T.; Ortiz, V. *Chem. Rev.* **2016**, *116* (11), 6488.
- (34) Glykos, N. M. *J. Comput. Chem.* **2006**, *27* (14), 1765.
- (35) Lange, O. F.; Grubmüller, H. *Proteins* **2006**, *62* (4), 1053.
- (36) Schoeler, C.; Malinowska, K. H.; Bernardi, R. C.; Milles, L. F.; Jobst, M. a.; Durner, E.; Ott, W.; Fried, D. B.; Bayer, E. a.; Schulten, K.; Gaub, H. E.; Nash, M. a. *Nat. Commun.* **2014**, *5*, 5635.



### 3.2 PUBLICATION P2: Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy

This article reports how protein receptors increase their mechanical stability upon binding of a ligand. We analyse the change in mechanical stability of a library of cohesins domains upon binding of a dockerin, because the mechanical stability of cohesin domains themselves is crucial to ensure cellulosome functionality in mechanically challenging environments. As target system, we investigated the structurally highly similar cohesins from *A. cellulolyticus*' main cellulosome scaffold ScaA and two mechanically stabilizing mutants from publication P1 through high-throughput AFM-based SMFS measurements.

To ensure comparability and avoid cantilever calibration errors, we employed the fast sample preparation protocol for the one-pot *in vitro* expression and surface attachment of multiple proteins in parallel from publication P1. This way we were able to probe nine cohesins - both without and with a dockerin bound - within 24 hours using a single cantilever, thus achieving unprecedented experimental throughput for absolute force comparisons. We found that all cohesins increase their mechanical stability, although the cohesins binding interface is located at the opposite site of the mechanical clamp motif. On average we determined an increase in most probable unfolding force of 95pN (+27%) upon binding of the dockerin.

Furthermore, we recorded dynamic force spectra for the unfolding of two cohesins and found that some cohesins undergo a transition from what seems to be a multitude of folds or unfolding pathways to a single stable fold or unfolding pathway upon binding of the dockerin.

# Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy

Tobias Verdorfer and Hermann E. Gaub

*published in*

Scientific Reports, 8, 9634, (2018)

*Reprinted from [114], with permission from Nature Publishing Group  
under a Creative Commons (CC BY 4.0) license*

# SCIENTIFIC REPORTS

## OPEN Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy

Received: 18 April 2018  
Accepted: 25 May 2018  
Published online: 25 June 2018

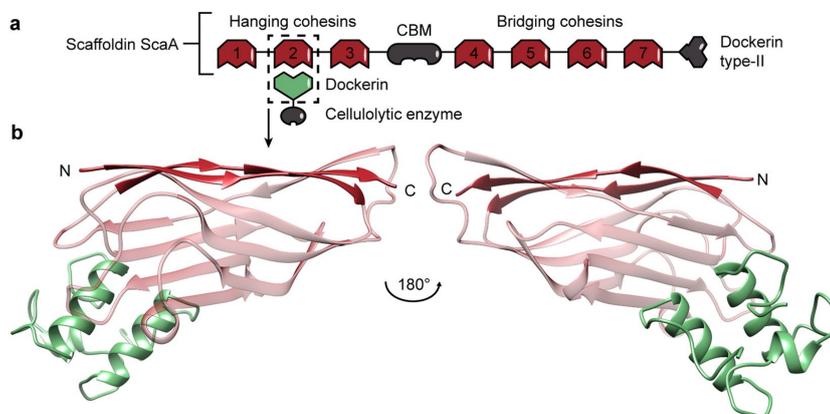
Tobias Verdorfer & Hermann E. Gaub

The cohesin-dockerin receptor-ligand family is the key element in the formation of multi-enzyme lignocellulose-digesting extracellular complexes called cellulosomes. Changes in a receptor protein upon binding of a ligand - commonly referred to as allostery - are not just essential for signalling, but may also alter the overall mechanical stability of a protein receptor. Here, we measured the change in mechanical stability of a library of cohesin receptor domains upon binding of their dockerin ligands in a multiplexed atomic force microscopy-based single-molecule force spectroscopy experiment. A parallelized, cell-free protein expression and immobilization protocol enables rapid mechanical phenotyping of an entire library of constructs with a single cantilever and thus ensures high throughput and precision. Our results show that dockerin binding increases the mechanical stability of every probed cohesin independently of its original folding strength. Furthermore, our results indicate that certain cohesins undergo a transition from a multitude of different folds or unfolding pathways to a single stable fold upon binding their ligand.

Cellulosomes are extracellular multi-enzyme complexes produced by many microorganisms for the efficient degradation of cellulose - nature's most abundant biopolymer. This degradation of plant cell-wall polysaccharides is accomplished via the spatial organization of a variety of cellulolytic enzymes through scaffolding proteins (i.e. scaffoldins) and boosted by its synergistic effects<sup>1,2</sup>. This complex network formation is driven by non-covalent, high-affinity receptor-ligand protein domains called cohesins (Coh) - which comprise the majority of scaffoldins - and dockerins (Doc) - which are typically connected to enzymatic domains<sup>3</sup>. The Coh-Doc complex has become a popular model system to study different aspects of protein-protein interactions, such as receptor-ligand binding specificities, affinities and strengths, dual binding conformations<sup>4-8</sup>, and the mechanical stability of cohesin domains - and thus the integrity of the cellulosomal scaffoldins themselves. In particular, it was shown that bridging cohesins, which are subjected to mechanical stress when the cell is anchored to its substrate, are able to withstand higher forces than hanging cohesins in order to remain folded and thus functional<sup>9-11</sup>. While these studies focussed exclusively on cohesins from the scaffoldin CipA of *Clostridium thermocellum* and the scaffoldin ScaA of *Acetivibrio cellulolyticus* in isolation, we now investigate the impact of dockerin binding on the mechanostability of cohesin folds.

The binding of a ligand can often change the fold and thereby the function of a receptor protein. For example, G protein-coupled receptors exhibit a wide variety of signaling behaviours in response to different ligands<sup>12</sup>. Ligand binding can also reduce the receptor's conformational folding space, as in the case of many intrinsically disordered proteins that only undergo folding upon binding to their physiological partner<sup>13</sup>. Furthermore, the binding of a protein ligand can result in significant improvement of a receptor's mechanical stability without introducing major structural changes, as reported by Cao *et al.*<sup>14</sup>. In agreement with this previous study, we discovered here that all probed cohesin domains markedly increase their force resilience upon binding of their dockerin ligand. This finding was enabled by combining several recent developments in atomic force microscopy-based (AFM) single-molecule force spectroscopy (SMFS), including site-specific protein immobilization, molecular fingerprint domains, highly specific and strong pulling handles, and *in vitro* transcription/

Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799, Munich, Germany. Correspondence and requests for materials should be addressed to T.V. (email: [tobias.verdorfer@physik.uni-muenchen.de](mailto:tobias.verdorfer@physik.uni-muenchen.de))



**Figure 1.** Schematic representation of scaffoldin ScaA and cohesin-dockerin binding geometry. (a) *A. cellulolyticus*' scaffoldin ScaA consists of seven type-I cohesins. Three mechanically weaker hanging cohesins (Coh1-Coh3) are located at the N-terminal end of the scaffoldin, while four mechanically stronger bridging cohesins (Coh4-Coh7) are located between a CBM and a type-II dockerin. The CBM and the type-II dockerin anchor the scaffoldin to the cellulolytic substrate and to the cell, respectively. The binding specificity of the cohesins is orthogonal to the type-II dockerin, thus they do not bind one another. Enzyme-bearing dockerins (green) can bind to each of the cohesins. ScaA's N-terminal glycoside hydrolase is not shown for simplicity. (b) Molecular representation of a typical cohesin-dockerin pair (PDB: 4DH2), as highlighted in Fig. 1(a). The cohesin is shown in red and the dockerin in green. The cohesin's binding interface is at the opposite site of its main structural element, the mechanical clamp motif, which consists of hydrogen bonds between parallel N- and C-terminal beta sheets.

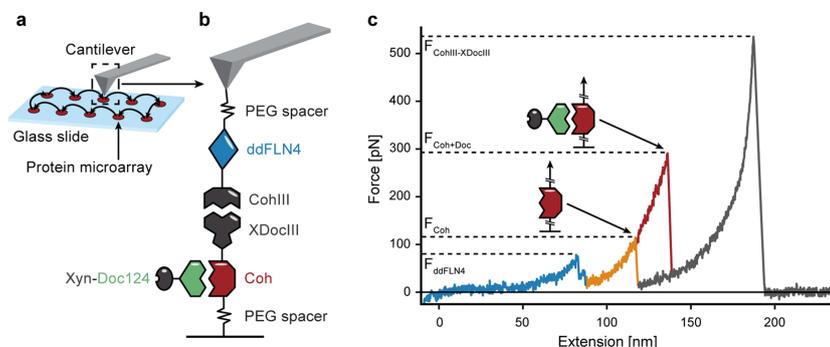
translation IVTT-based microarray-format sample preparation<sup>10,15,16</sup>. Taken together, these techniques allow for a high-throughput characterization of the mechanical stability of a library of receptors - without and with a ligand bound - in a single automated AFM-SMFS experiment.

Our study focuses on all seven cohesins from *A. cellulolyticus*' scaffoldin ScaA (Fig. 1(a)) and two cohesin mutants from our previous work<sup>10</sup>. Scaffoldin ScaA contains three major domains: a series of three hanging cohesins, a central carbohydrate-binding module (CBM) that anchors the cell to its substrate, and a series of four bridging cohesins. Bridging cohesins are mechanically stronger than hanging cohesins, although these two classes both bind *A. cellulolyticus*' type-I dockerins<sup>17</sup>. Additionally, we tested two mutants of ScaA's mechanically weakest cohesin Coh1 from our previous work, where we identified minimal mutations that significantly increased the mechanical stability of Coh1. Mutant T107S displayed an atypical bimodal unfolding force distribution and we hypothesized that either strongly differing fold conformations or unfolding pathways could cause such a behaviour. Mutant T107S was included in this study to clarify if it can bind dockerins and whether binding of a dockerin would stabilize its fold or alter its unfolding behaviour. Furthermore, we included mutant GGS, which is a triple mutant of Coh1 with the mutations A105G, P106G, and T107S. This mutant showed the largest increase in mechanical stability when compared to wild-type Coh1.

The high mechanical resilience of cellulosomal cohesins is predominantly attributed to the so-called mechanical clamp motif: a set of backbone hydrogen bonds formed between the N- and C-terminal  $\beta$ -strands of the cohesin protein domains (Fig. 1(b)). Previous studies that pulled cohesin from its termini - including all-atom steered molecular dynamics simulations - have identified the mechanical clamp motif as their main structural element<sup>9,10</sup>. Since the cohesin's dockerin binding interface is located at the opposite side of the protein, a mechanically stabilizing effect of dockerin binding would thus have to be propagated through the cohesin fold to affect its unfolding energy landscape.

## Results and Discussion

**Experimental setup.** A library of fusion proteins containing the cohesins of interest is expressed and immobilized at spatially separated positions in a microarray format on glass slide using a cell-free, one-pot protein expression and pulldown reaction<sup>10</sup>. In short, a silicone mask is placed on a glass slide to compartmentalize the surface into several microwells with a diameter of about 5 mm. All microwells are functionalized with reactive PEG-linkers and filled with a one-pot reaction mix consisting of the *in vitro* transcription/translation (IVTT) system, the plasmid DNA encoding for the fusion proteins, and the enzyme facilitating the site-specific protein pulldown (phosphopantetheinyl transferase). In a single incubation step, this mixture results in cell-free protein synthesis, and simultaneous covalent ligation of the proteins onto the surface. The spatially separated protein spots can be probed in series using a single functionalized cantilever (see Figure 2(a)). This parallelized sample preparation method thus allows for quick mechanical phenotyping of protein libraries through AFM-SMFS.



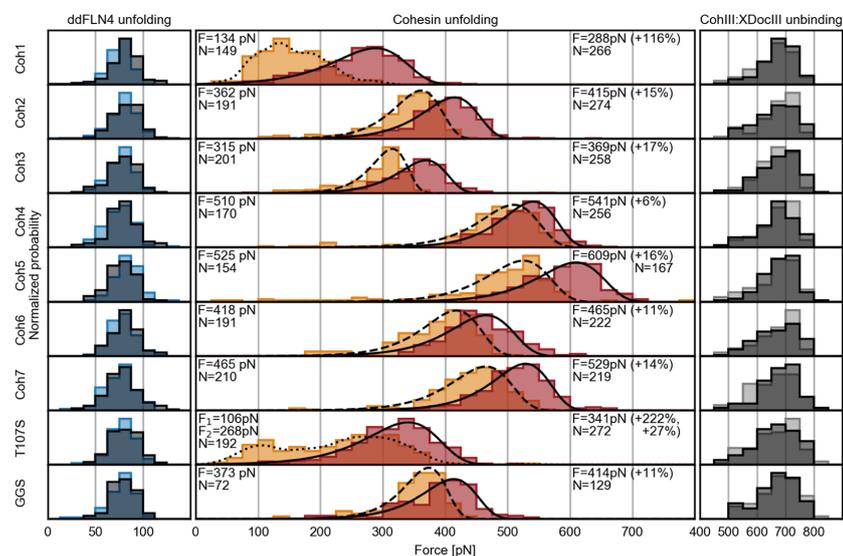
**Figure 2.** Experimental design and exemplary force curve. (a) One AFM cantilever can probe a library of different protein constructs, which are immobilized on a glass slide in a microarray format. This protocol allows for high experimental throughput and precise relative force comparability. (b) A fusion protein consisting of a ddFLN4 fingerprint domain and a CohIII is immobilized on the cantilever. Fusion proteins, each consisting of a cohesin of interest (Coh) as well as an XDocIII, are expressed and immobilized in a microarray format on a glass slide. The CohIII-XDocIII receptor-ligand pair serves as a highly specific pulling handle. At a later stage in the experiment, Xyn-Doc124 is added to the experimental buffer solution and can then bind to each Coh. (b) Overlay of two exemplary force-distance curves when retracting the cantilever with constant speed. First, the poly-protein stretches and the ddFLN4 fingerprint domain unfolds in two steps and at a relatively low force  $F_{FP}$  (blue). The poly-protein is stretched further and the cohesin domain unfolds at  $F_{Coh}$  (orange). In the presence of Xyn-Doc124, the cohesin unfolds at a higher unfolding force  $F_{Coh+Doc}$  (red). Finally, the CohIII-XDocIII pulling handle unbinds at  $F_{CohIII-XdocIII}$  (grey) and the force ultimately drops back to zero.

Figure 2(b) shows the molecular pulling configuration. We chose the type-III cohesin-dockerin complex (CohIII-XDocIII) from *Ruminococcus flavefaciens* as a pulling handle due to its high specificity, long-term stability and high unbinding force<sup>6</sup>. The CohIII-XDocIII interaction is not expected to bind the type-I cohesins and dockerins from *A. cellulolyticus*, since the interaction between cohesins and dockerins was shown to be highly type- and species-specific<sup>17</sup>. The fusion proteins immobilized in the individual microarray spots on the glass slide each consist of a cohesin of interest (labeled Coh in Figure 2(b)) and an XDocIII. On the cantilever side we immobilized CohIII-ddFLN4 fusion proteins. ddFLN4 is the 4th immunoglobulin rod filamin domain from *Dictyostelium discoideum* and serves as a molecular fingerprint<sup>18,19</sup>. Its molecular unfolding pattern (rupture force and contour length increment) is used to identify traces with single, specific interactions. As a binding partner for the cohesins under investigation, we chose the dockerin from *A. cellulolyticus*' enzyme-bearing Cel124A, as it has been shown to bind ScaA's cohesins<sup>17</sup>. Subsequently, we designed a xylanase-dockerin124 (Xyn-Doc124) fusion construct derived from it. After all cohesins are probed without a dockerin bound, Xyn-Doc124 was added to the measurement buffer. We used a concentration of 5  $\mu$ M, which is well above the typical affinity constant of a type-I cohesin-dockerin pair of  $\sim 10$ -11 M, thus ensuring that the vast majority of cohesin domains should have a dockerin bound during measurement<sup>4</sup>.

With this experimental design, all cohesins of interest are probed using a single cantilever *via* the CohIII-XDocIII pulling handle in a single high-throughput experiment. The cantilever automatically cycles through the protein microarray, acquiring hundreds of force-distance traces for each Coh before moving to the next position. When the cantilever approaches the surface in any of the microarray protein spots, a CohIII-XDocIII receptor ligand interaction can form. The cantilever retracts with constant speed, stretching the polypeptide chain. A typical AFM SMFS force-distance trace is shown in Fig. 2(c). First, the ddFLN4 domain unfolds in a recognizable two-step pattern, followed by the unfolding of the cohesin domain and the final unbinding of the CohIII-XDocIII pulling handle. It is possible that the CohIII unfolds as the CohIII-XDocIII bond ruptures, but our constant yield of force curves over 24 hours and several thousand force traces is a clear indication that CohIII either stays intact during the course of the experiment or quickly refolds after unfolding. The rupture force of the individual peaks ( $F_{ddFLN4}$ ,  $F_{Coh}$  and  $F_{CohIII-XdocIII}$ , respectively) is extracted from the data and analysed further. After collecting sufficient unfolding data of the bare cohesins, the dockerin is added to the experimental buffer and the AFM continuously probes the ligand-bound cohesins. When Xyn-Doc124 is present in the measurement buffer, a higher cohesin unfolding peak is typically recorded ( $F_{Coh+Doc} > F_{Coh}$ ).

By using a single cantilever through all conditions, calibration errors of up to  $\sim 15\%$  that normally result from individual AFM-based SMFS measurements can be circumvented<sup>20</sup>. We therefore obtain comparable absolute force data of all probed cohesins both with and without the dockerin bound.

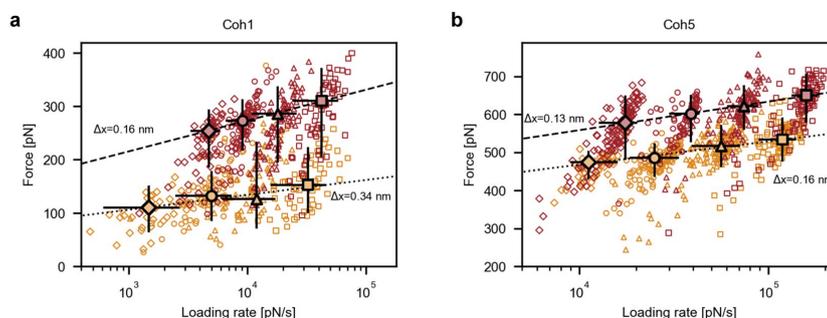
**Cohesin mechanostability increases upon binding of dockerin.** Figure 3 shows the unfolding force distributions of the ddFLN4 fingerprint domain (left column, blue) and of the cohesins under investigation (middle column, red), as well as the unbinding force distribution of the orthogonal CohIII-XDocIII pulling handle (right column, grey). Data collected without and with Xyn-Doc124 in solution are represented in dark and bright



**Figure 3.** Unfolding and rupture force histograms without and with Xyn-Doc124 in solution. Both the unfolding force histograms of the ddFLN4 fingerprint (left column, bright and dark blue without and with dockerin in solution, respectively) and the unbinding force histograms of the CohIII-XDdocIII pulling handle (right column, bright and dark grey without and with dockerin in solution, respectively) are independent of the cohesin under investigation and are unaffected by the presence or absence of Xyn-Doc124 during the measurement. All cohesins under investigation show their expected and previously known unfolding force distributions without Xyn-Doc124 in solution (middle column, orange)<sup>10</sup>. All but Coh1 and T107S can be fitted adequately using the Bell-Evans model to obtain their most probable rupture forces (dashed lines). The rupture force histograms of Coh1 and T107S were approximated using KDEs (dotted lines) to obtain the most probable rupture forces. In the presence of Xyn-Doc124, the unfolding force distributions of all cohesins are shifted towards higher forces (middle column, red) and can be fitted using the Bell-Evans model (solid lines). Forces  $F$  on the left and right side of the middle column represent the most probable rupture forces of the cohesins without and with Xyn-Doc124 in solution, respectively, and their relative change in percent. Numbers  $N$  on the left and on the right side of the middle column represent the number of cohesin unfolding force data points in the histograms without and with dockerin in solution, respectively. We found that changing the bin size within a range of 10 pN to 50 pN did not notably alter the fits of the Bell-Evans model to the cohesin unfolding force histograms or the obtained most probable unfolding forces. We therefore chose the same bin size for all constructs within a column of histograms (25 pN for all ddFLN4 and cohesin unfolding events, and 50 pN for all CohIII-XDdocIII unbinding events), to provide good comparability by eye. All data were recorded using a single cantilever with a spring constant of 143 pN/nm at a retraction speed of 1600 nm/s in a 24-hour automated SMFS experiment. For more information see Table S1.

colors, respectively. The force histograms of both the ddFLN4 fingerprint and of the CohIII-XDdocIII pulling handle agree with previously reported literature values<sup>6,18</sup>. Furthermore, they are independent of the cohesin under investigation and are unaffected by the presence or absence of Xyn-Doc124 during the measurement (see Table S1(a,c)). This confirms precise relative force comparability among the cohesins under investigation. All cohesins show their expected unfolding force distributions without Xyn-Doc124 in the measurement buffer, which we already determined in previous work<sup>10</sup>. As expected, all cohesin unfolding force distributions except those of Coh1 and T107S can be fitted adequately using the conventional two-state Bell-Evans model to obtain their most probable rupture forces (dashed lines)<sup>21,22</sup>. The rupture force histograms of Coh1 and T107S are more complex, indicating a multi-barrier unfolding energy landscape or multiple folds, and were therefore processed with kernel density estimation (KDE) to obtain most probable rupture forces from them (dotted lines). This atypical unfolding behavior might be explained by a distinctly flexible region critical for cohesin stability, for a detailed discussion see our previous work<sup>10</sup>. Since mutant T107S exhibits a bimodal unfolding force distribution, two most probable rupture forces  $F_1$  and  $F_2$  were extracted.

Upon addition of Xyn-Doc124 to the measurement buffer, all cohesins unfolded at notably higher forces (see Table S1(b)) and now all cohesin unfolding force distributions, including Coh1's and T107S', could be fitted with the Bell-Evans model (solid lines). Coh4 shows the smallest change in most probable rupture force with an increase of 31 pN (+6%). Cohesins number 2, 3, 6, 7 and mutant GGS all show an increase of 40 pN (+11%) to 84 pN (+17%). Interestingly, both Coh1 and mutant T107S, which displayed an atypical unfolding force distribution



**Figure 4.** Dynamic force spectra for unfolding of Coh1 and Coh5, both without and with dockerin bound (orange and red, respectively). Diamonds, circles, triangles and squares represent pulling speeds of 400 nm/s, 800 nm/s, 1600 nm/s and 3200 nm/s, respectively. Large markers represent the most probable rupture force/loading rates of each speed. Error bars represent the full widths at half maxima. Fits of the Bell-Evans model through the most probable rupture force/loading rates of each speed are represented by dotted and broken lines (no dockerin and dockerin bound, respectively) (a) Fitting the Bell-Evans model to the most probable unfolding events of Coh1 of each pulling speed yields the distance to the transition state ( $\pm$ SD) of  $\Delta x = 0.34 \pm 0.11$  nm and zero-force off rate  $k_0 = (1.3 \pm 4.1) 10^{-2} \text{ s}^{-1}$  for the cohesin alone, and  $\Delta x = 0.16 \pm 0.01$  nm and  $k_0 = (8.6 \pm 4.0) 10^{-3} \text{ s}^{-1}$  with a dockerin bound. (b) The Bell-Evans fit to the most probable unfolding events of Coh5 of each pulling speed yields distances to the transition state ( $\pm$ SD) of  $\Delta x = 0.16 \pm 0.02$  nm and zero-force off rate  $k_0 = (11.7 \pm 9.3) 10^{-6} \text{ s}^{-1}$ , for the cohesin alone, and  $\Delta x = 0.13 \pm 0.01$  nm and  $k_0 = (9.1 \pm 2.1) 10^{-6} \text{ s}^{-1}$ , with a dockerin bound. All data were recorded using a single cantilever with a spring constant of 117 pN/nm during a 24 hour automated SMFS experiment.

without dockerin, exhibit the largest increase in unfolding force upon addition of the dockerin, with  $\Delta F_{\text{Coh1}} = 154$  pN (+116%), and  $\Delta F_{\text{T107S,1}} = 235$  pN (+222%) and  $\Delta F_{\text{T107S,2}} = 73$  pN (+27%). This remarkable mechanical stabilization and the recovery of the Bell-Evans shape in their unfolding force distributions hint that the folds of Coh1 and T107S stabilize upon binding of the dockerin and that unfolding is now dominated by a single barrier. The average increase in most probable unfolding force of all cohesins under investigation at a loading rate which corresponds to an AFM pulling speed of 1600 nm/s was 95 pN (+27%), while both the ddFLN4 fingerprint unfolding and the CohIII-XDocIII unbinding showed no notable change (see Table S1).

**Negative control.** To rule out a stabilizing effect caused by the Xyn domain or any other unknown effect caused by the addition of 5  $\mu\text{M}$  protein on the cohesins' unfolding behaviour, we added 5  $\mu\text{M}$  xylanase-dockerinS (Xyn-DocS) instead of Xyn-Doc124 in a second measurement. This measurement served as a negative control, since DocS from *C. thermocellum*'s exocellulase Cel48S is known to not bind any of the cohesins under investigation<sup>17</sup>. We tested the change in unfolding force of the four cohesins which showed the strongest change under addition of Xyn-Doc124 (Coh1, Coh3, Coh5 and T107S) and found no notable change in unfolding behaviour for any of them (see Figure S2 and Table S3). This result confirms that the increase in mechanical stability upon addition of Xyn-Doc124 is in fact induced by the specific binding of Doc124 to ScaA's cohesins, and not the presence of Xyn.

**Characterization of the free-energy landscape of Coh1 and Coh5.** To better characterize the underlying conformational changes resulting in the increased mechanostability of the cohesins upon binding of the dockerins, we recorded dynamic force spectra of Coh1 and Coh5. The forces needed to unfold the two cohesins - without and with dockerin bound - were measured as a function of the loading rate by varying the AFM pulling speed between 400 nm/s, 800 nm/s, 1600 nm/s and 3200 nm/s.

We chose Coh1 because of its transition from an atypical unfolding force distribution to one which can be described by the Bell-Evans theory for single-barrier unfolding events. Additionally, Coh1 shows the largest absolute increase in most probable unfolding force of 154 pN. Coh5 was chosen because it is mechanically the strongest of all cohesins under investigation, both with and without a dockerin bound, and because it exhibits the second largest absolute increase in unfolding force of 84 pN. Figure 4 shows the dynamic force spectra for Coh1 and Coh5, without and with dockerin bound, in dark and bright red, respectively. Although Coh1's unfolding force distribution with no dockerin bound cannot be described by a Bell-Evans distribution, its dynamic force spectrum shows a linear dependence of the unfolding force on the logarithm of the force loading rate, which is predicted by the Bell-Evans model for dynamic force spectroscopy. We therefore chose to fit all dynamic force spectra, including Coh1's, with the Bell-Evans model to obtain the parameters  $\Delta x$ , the distance to transition state in the protein unfolding energy landscape, and  $k_0$ , the zero-force off rate, and evaluate the change in the obtained kinetic parameters upon dockerin binding.

The fit of the Bell-Evans model to Coh1's unfolding events without dockerin reveals an atypically large distance to the transition state ( $\pm$ SD) of  $\Delta x = 0.34 \pm 0.11$  nm compared to previously measured values for cohesins

of 0.11 nm to 0.17 nm<sup>5,9,10</sup>. The binding of the dockerin reduces this value to  $\Delta x = 0.16 \pm 0.01$  nm. This is an indication that Coh1 undergoes a transition from one or more relatively weak fold conformations to a more stable folding state. The distance to the transition state of the strongest probed cohesin Coh5 slightly decreases from  $\Delta x = 0.16 \pm 0.02$  nm to  $\Delta x = 0.13 \pm 0.01$  nm, indicating a minor to no increase of the rigidity of its fold upon ligand binding.

### Conclusions

In summary, we characterized the change in mechanical stability of all seven cohesins from *A. cellulolyticus*' primary scaffoldin ScaA and two mutants of Coh1, utilizing a parallelized and high-throughput AFM-based SMFS protocol. All cohesins under investigation remarkably increased their mechanical stability upon binding of the dockerin, even though ligand binding takes place at the opposite site of the cohesins main structural element - i.e. the mechanical clamp motif - and the physical opening of the fold. While a recent study by Galera-Prat *et al.* found that two cohesins from *C. thermocellum*'s scaffoldin CipA did not change their mechanical resilience in the presence of a ligand<sup>11</sup>, we determined an average increase in most probable unfolding force for the cohesins of *A. cellulolyticus*' scaffoldin ScaA of 95 pN (+27%) upon dockerin binding at a loading rate which corresponds to an AFM pulling speed of 1600 nm/s. Additionally, the change in unfolding force distributions of Coh1 and T107S indicates that both cohesins were stabilized by dockerin binding. We hypothesize that Coh1 and T107S transition either from an ensemble of folding states or *via* several unfolding pathways to the unfolded state, unless stabilized in a single fold through binding of the dockerin, resulting in an unfolding behaviour dominated by a single barrier. This hypothesis is further strengthened by the observation that Coh1's distance to the transition state decreases when its ligand is bound. Moreover, this suggests that Coh1 transitions through several fold conformations - similar in fold energy and separated only by minor barriers - during unfolding in the absence of dockerin.

We have employed diverse experimental techniques in this study, including the fast and parallelized one-pot *in vitro* protein expression and site-specific pulldown, highly specific pulling handles and molecular fingerprints, and parallel measurement of a protein library with a single cantilever. The combination of these experimental protocols allowed us to characterize the mechanical stability of nine receptors - with and without a ligand bound - within one 24-hour AFM-based SMFS experiment. Combined with recent developments in AFM-based SMFS, which have both accelerated data acquisition and improved data quality, the allosteric effects on protein mechanics induced by receptor-ligand binding can be screened more rapidly and accurately<sup>15,16,23-26</sup>.

### Materials and Methods

All reagents were at least of analytical purity grade and all buffers were filtered using a 0.2  $\mu$ m polyethersulfone membrane filter prior to use. All incubation steps were done at room temperature, if not otherwise stated. All protocols follow Verdorfer *et al.*<sup>10</sup>. All data and constructs are available upon reasonable request.

**Gene construction, protein expression and purification.** All genes were codon optimized for *E. coli*, synthesized and cloned into pET28a vectors. Plasmid DNA for cohesins mutants T107S and GGS was constructed from Coh1's plasmid DNA using individually designed primers. All plasmids used in *in vitro* protein expression were amplified in DH5- $\alpha$  cells, purified using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany), eluted with ultrapure water and stored at  $-20^{\circ}\text{C}$ . All sequences were finally checked by DNA sequencing (Eurofins Genomics GmbH, Ebersberg, Germany).

The pulling handle and fingerprint protein CohIII-ddFLN4-HIS-ybBR, as well as both Xylanase-Dockerin (xylanase is typically used to enhance dockerin expression and folding<sup>27</sup>) constructs were expressed in *E. coli* NiCo21 (DE3) cells (New England Biolabs, MA, USA) and purified using a Ni-NTA column for HIS-Tag purification. Both dockerins (Doc124 and DocS) could not to be expressed with their natural cellulases attached. The proteins were concentrated and exchanged into measurement buffer (TBS - Ca: 25 mM Tris, 72 mM NaCl, 1 mM CaCl<sub>2</sub>, pH 7.2) using desalting columns. They were frozen with 25% v/v glycerol in liquid nitrogen to be stored at  $-80^{\circ}\text{C}$  until used in experiments. All protein sequences can be found in the SI.

**AFM Sample preparation.** Both the AFM cantilevers and the microscope slides were silanized using (3-aminopropyl)-dimethyl-ethoxysilane. A multiwell mask was attached to the glass slide to allow compartmentalization of the surface in spots with a diameter of  $\sim 500$   $\mu$ m. Both the cantilevers and the individual surfaces in the wells were incubated with 20 mM NHS-PEG-Maleimide (5 kDa) in 100 mM HEPES buffer pH 7.4 for 45 minutes. After rinsing with ultrapure water, both the cantilevers and the surfaces were incubated with 1 mM Coenzyme A (CoA), which bonds to the PEG's maleimide group, in a 1 mM sodium phosphate pH 7.2, 50 mM NaCl, 10 mM EDTA buffer for 2 hours. After a final ultrapure water rinse the cantilevers were incubated with 40  $\mu$ M Coh3-ddFLN4-HIS-ybBR and 5  $\mu$ M phosphopantetheinyl transferase (Sfp)<sup>28</sup> for 2 hours in magnesium chloride supplemented measurement buffer (TBS - Ca: 25 mM Tris, 72 mM NaCl, 1 mM CaCl<sub>2</sub>, 20 mM MgCl<sub>2</sub>, pH 7.2). The Sfp covalently bonds the Coh3-ddFLN4-HIS-ybBR to the PEG-CoA linkers on the cantilever. The cantilevers were rinsed extensively with measurement buffer and finally stored in it until use in measurement.

***In vitro* expression and protein pulldown.** The IVTT reaction mixes consisting of the PURExpress<sup>®</sup> kit, 5  $\mu$ M Sfp<sup>28</sup>, 0.8 U/ $\mu$ l RNase inhibitor, 10 ng/ $\mu$ l Plasmid-DNA (encoding for ybBR-Coh-XDocIII) and 0.05% v/v Triton X-100 were transferred to the wells and incubated at 37  $^{\circ}\text{C}$  for 3 h. The proteins were both expressed, and bound to the PEG-CoA linkers on the surface by the Sfp *via* their ybBR-tag during this time. The individual wells were finally rinsed extensively using measurement buffer and the multiwell mask was removed.

**AFM SMFS measurements.** A custom-build AFM, specialized for multispot SMFS, was used for all measurements. The cantilever was aligned to all individual protein spots and the positions were stored in the control software. All single-speed SMFS measurements were done with a constant pulling velocity of 1600 nm/s. The

dynamic force spectra were recorded at a pulling speed of 400 nm/s, 800 nm/s, 1600 nm/s, and 3200 nm/s. After 2000 approach- and retract-cycles in one protein spot the AFM head was automatically moved to the next one. The measurement buffer on the sample was exchanged after 6 to 12 hours with measurement buffer containing 5  $\mu$ M of Xylanase-Dockerin; Doc124 from *A. cellulolyticus* or DocS (taken from exocellulase Cel48S) from *C. thermocellum* for main measurement and dynamic force spectra or negative control, respectively.

**AFM SMFS data Analysis.** Data analysis was carried out following previous work<sup>10,19</sup>. Laser spot drift on the cantilever relative to the calibration curve can be significant in 24-hour long measurement sessions. It was therefore corrected via the baseline noise (determined as the last 5% of data points for each curve) for all curves and smoothed with a moving median. The inverse optical lever sensitivity (InvOLS) for each curve was linearly corrected relative to the InvOLS value of the calibration curve according to the baseline noise. All curves showing a ddFLN4 and cohesin contour length increment (ddFLN4: 34 nm<sup>18,19</sup>, cohesin: 45 nm) were used to assemble unfolding force histograms, which were then fitted either following the Bell-Evans model<sup>22,29</sup> or using kernel density estimates. Both methods were used to obtain most probable rupture forces.

Bell-Evans probability density function at given loading rate  $r$  with fit parameters distance to the transition state  $\Delta x$  and natural off-rate  $k_0$ :

$$p(F) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 k_B T}{r \Delta x} (e^{\frac{\Delta x F}{k_B T}} - 1)\right]$$

The Bell-Evans model predicts a linear dependence between the most probable rupture force  $\langle F \rangle$  and the logarithm of the force loading rate  $r$  in dynamic force spectra:

$$\langle F(r) \rangle = \left(\frac{k_B T}{\Delta x}\right) \ln\left(\frac{r \Delta x}{k_0 k_B T}\right)$$

The standard deviation for the fitted parameters  $\Delta x$  and  $k_0$  were obtained by taking the square root of the diagonal entries of the covariance matrix from the fitting algorithm. We used the Levenberg-Marquardt algorithm from the Scipy python library<sup>30,31</sup>.

## References

- Bayer, E. A., Morag, E. & Lamed, R. The cellulosome—a treasure-trove for biotechnology. *Trends Biotechnol.* **12**, 379–386 (1994).
- Bayer, E. a., Lamed, R., White, B. a. & Flint, H. J. From cellulosomes to cellulosomes. *Chem. Rec.* **8**, 364–377 (2008).
- Carvalho, A. L. *et al.* Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. *Proceedings of the National Academy of Sciences.* **100**, 13809–13814 (2003).
- Mechaly, A. *et al.* Cohesin-Dockerin Interaction in Cellulosome Assembly: A Single Hydroxyl Group of a Dockerin Domain Distinguishes Between Nonrecognition and High Affinity Recognition. *J. Biol. Chem.* **276**, 9883–9888 (2001).
- Stahl, S. W. *et al.* Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. USA* **109**, 20431–20436 (2012).
- Schoeler, C. *et al.* Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.* **5**, 5635 (2014).
- Jobst, M. A. *et al.* Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *Elife.* **4**, 1–19 (2015).
- Bule, P., Pires, V. M., Fontes, C. M. & Alves, V. D. Cellulosome assembly: paradigms are meant to be broken! *Curr. Opin. Struct. Biol.* **49**, 154–161 (2018).
- Valbuena, A. *et al.* On the remarkable mechanostability of scaffolds and the mechanical clamp motif. *Proc. Natl. Acad. Sci. USA* **106**, 13791–13796 (2009).
- Verdorfer, T. *et al.* Combining *In Vitro* and *In Silico* Single Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J. Am. Chem. Soc.* <https://doi.org/10.1021/jacs.7b07574> (2017).
- Galera-Prat, A., Morais, S., Vazana, Y., Bayer, E. A. & Carrión-Vázquez, M. The cohesin module is a major determinant of cellulosome mechanical stability. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.RA117.000644> (2018).
- Kim, T. H. *et al.* The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *J. Am. Chem. Soc.* **135**, 9465–9474 (2013).
- Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38 (2009).
- Cao, Y., Yoo, T., Zhuang, S. & Li, H. Protein-protein interaction regulates proteins' mechanical stability. *J. Mol. Biol.* **378**, 1132–1141 (2008).
- Otten, M. *et al.* From genes to protein mechanics on a chip. *Nat. Methods*, 1–6 (2014).
- Durner, E., Ott, W., Nash, M. A. & Gaub, H. E. Post-Translational Sortase-Mediated Attachment of High-Strength Force Spectroscopy Handles. *ACS Omega.* **2**, 3064–3069 (2017).
- Hamberg, Y. *et al.* Elaborate cellulosome architecture of *Acetivibrio cellulolyticus* revealed by selective screening of cohesin–dockerin interactions. *PeerJ.* **2**, e636 (2014).
- Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A. A. & Rief, M. A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat. Struct. Mol. Biol.* **11**, 81–85 (2004).
- Milles, L. F., Bayer, E. A., Nash, M. A. & Gaub, H. E. Mechanical Stability of a High-Affinity Toxin Anchor from the Pathogen *Clostridium perfringens*. *J. Phys. Chem. B, acs.jpcc.6b09593* (2016).
- Brand, U. *et al.* Comparing AFM cantilever stiffness measured using the thermal vibration and the improved thermal vibration methods with that of an SI traceable method based on MEMS. *Measurement Science and Technology.* **28**, 034010 (2017).
- Bell, G. I. Models for the specific adhesion of cells to cells. *Science.* **200**, 618–627 (1978).
- Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. *Biophys. J.* **72**, 1541–1555 (1997).
- Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M. & Scheuring, S. High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations. *Science.* **342**, 741–743 (2013).
- Scholl, Z. N. & Marszalek, P. E. Improving single molecule force spectroscopy through automated real-time data collection and quantification of experimental conditions. *Ultramicroscopy.* **136**, 7–14 (2014).
- Walder, R. *et al.* Rapid Characterization of a Mechanically Labile  $\alpha$ -Helical Protein Enabled by Efficient Site-Specific Bioconjugation. *J. Am. Chem. Soc.* **139**, 9867–9875 (2017).
- Ott, W. *et al.* Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano.* **11**, 6346–6354 (2017).

27. Haimovitz, R. *et al.* Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. *Proteomics*. **8**, 968–979 (2008).
28. Yin, J., Lin, A. J., Golan, D. E. & Walsh, C. T. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.* **1**, 280–285 (2006).
29. Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.* **72**, 1568–1581 (1997).
30. Oliphant, T. E. Python for Scientific Computing. *Computing in Science Engineering*. **9**, 10–20 (2007).
31. Jones, E., Oliphant, T. & Peterson, P., SciPy: Open source scientific tools for Python (2001–), (available at <http://www.scipy.org>).

### Acknowledgements

Support for this work was provided by the ERC Advanced Grant CelluFuel. We thank E. A. Bayer and M. Carrión-Vázquez for helpful discussions, L. F. Milles for providing the force curve analysis software, as well as E. Durner and M. A. Jobst for the development of the instrument control software, and W. Ott and T. Nicolaus for cloning- and laboratory assistance.

### Author Contributions

T.V. experiment design, protein design and expression, sample preparation, measurements, data analysis, writing of manuscript, preparation of figures; H.E.G. experiment design, writing of manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-27085-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Supporting Information for

Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy

Tobias Verdorfer\*<sup>1</sup> and Hermann E. Gaub<sup>1</sup>

<sup>1</sup> Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany.

\*Correspondence and requests for materials should be addressed to TV (email: tobias.verdorfer@physik.uni-muenchen.de).

KEYWORDS Single molecule force spectroscopy, AFM, cohesin-dockerin, cellulosome, mechanical stability, receptor-ligand, allostery

(a)

	ddFLN4 unfolding forces F [pN]						
	most probable				mean		
	no doc	doc	ΔF		no doc	doc	ΔF
Cohesin 1	82	81	-2	-2.1%	80	76	-4 -5.3%
Cohesin 2	86	82	-4	-4.9%	82	79	-2 -3.0%
Cohesin 3	82	78	-3	-4.1%	78	77	-1 -1.8%
Cohesin 4	81	78	-3	-4.0%	79	75	-4 -5.0%
Cohesin 5	81	78	-2	-3.1%	78	83	5 6.4%
Cohesin 6	82	78	-4	-4.9%	82	78	-3 -3.9%
Cohesin 7	77	77	0	-0.3%	79	75	-3 -4.1%
Cohesin T107S	77	82	5	6.1%	80	79	-2 -2.0%
Cohesin GGS	77	81	5	6.1%	79	77	-2 -2.8%
<b>average</b>	<b>81</b>	<b>79</b>	<b>-1</b>	<b>-1.4%</b>	<b>80</b>	<b>78</b>	<b>-2 -2.4%</b>

(d)

	Number of curves	
	no doc	doc
Cohesin 1	149	266
Cohesin 2	191	274
Cohesin 3	201	258
Cohesin 4	170	256
Cohesin 5	154	167
Cohesin 6	191	222
Cohesin 7	210	219
Cohesin T107S	192	272
Cohesin GGS	72	129
<b>sum</b>	<b>1530</b>	<b>2063</b>

(b)

	Cohesin unfolding forces F [pN]										
	most probable				mean						
	no doc	doc	ΔF		no doc	doc	ΔF				
Cohesin 1	134	288	154	<b>115.6%</b>	153	262	109 <b>71.5%</b>				
Cohesin 2	362	415	53	<b>14.8%</b>	336	395	59 <b>17.5%</b>				
Cohesin 3	315	369	54	<b>17.2%</b>	291	350	59 <b>20.1%</b>				
Cohesin 4	510	541	32	<b>6.2%</b>	478	522	43 <b>9.1%</b>				
Cohesin 5	525	609	84	<b>16.1%</b>	474	584	109 <b>23.0%</b>				
Cohesin 6	418	465	48	<b>11.4%</b>	387	455	68 <b>17.7%</b>				
Cohesin 7	465	529	64	<b>13.7%</b>	440	512	72 <b>16.3%</b>				
Cohesin T107S	106	268	341	235	73	<b>221.7%</b>	<b>27.1%</b>	226	319	93	<b>41.0%</b>
Cohesin GGS	373	414	40	<b>10.7%</b>	350	391	41 <b>11.6%</b>				
<b>average</b>	<b>348</b>	<b>441</b>	<b>94</b>	<b>27.0%</b>	<b>348</b>	<b>421</b>	<b>72</b>	<b>20.8%</b>			

(e)

lever spring constant: 143pN/nm

time w/o doc: ~12hrs  
time w/ doc: ~12hrs  
total exp time: ~24hrs

(c)

	CohIII:XDociII unbinding forces F [pN]						
	most probable				mean		
	no doc	doc	ΔF		no doc	doc	ΔF
Cohesin 1	693	674	-18	-2.6%	675	659	-16 -2.3%
Cohesin 2	693	700	8	1.1%	666	669	2 0.4%
Cohesin 3	702	694	-8	-1.2%	677	666	-11 -1.6%
Cohesin 4	685	701	16	2.3%	669	674	5 0.8%
Cohesin 5	703	700	-3	-0.5%	680	669	-11 -1.6%
Cohesin 6	715	710	-4	-0.6%	670	665	-4 -0.7%
Cohesin 7	698	691	-7	-1.0%	678	658	-20 -3.0%
Cohesin T107S	700	702	2	0.3%	665	665	-1 -0.1%
Cohesin GGS	683	727	44	6.4%	667	676	9 1.3%
<b>average</b>	<b>697</b>	<b>700</b>	<b>3</b>	<b>0.5%</b>	<b>672</b>	<b>667</b>	<b>-5 -0.8%</b>

Table S1: Summary of all unfolding force data from main experiment. (a)-(c) Most probable and mean rupture for the ddFLN4 fingerprint, all cohesins of interest and the CohIII-XDocIII pulling handle. Neither the ddFLN4's nor the CohIII-XDocIII's most probable or mean unfolding forces show a notable change upon addition of the dockerin Doc124, while all cohesins show a considerable increase. (d) Number of force curves of all molecular constructs. (e) General experiment information.

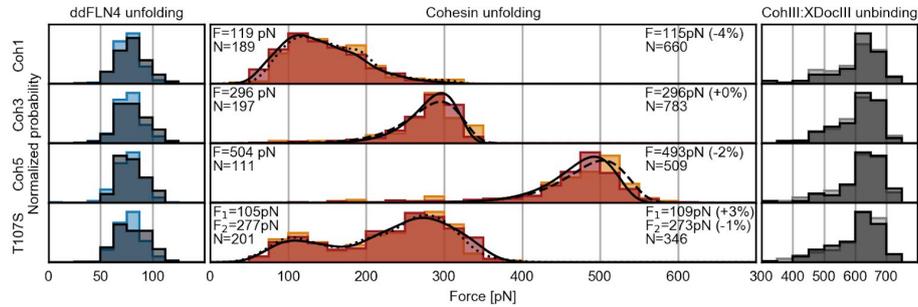


Figure S2: Negative control: Unfolding and rupture force histograms without and with Xyn-DocS in solution. Tested were the four cohesins with the strongest change unfolding force under addition of Xyn-Doc124: Coh1, Coh3, Coh5, and T107S. Both the unfolding force histograms of the fingerprint (left column, bright and dark blue without and with dockerin in solution, respectively) and the unbinding force histograms of the CohIII-XDdocIII pulling handle (right column, bright and dark grey without and with dockerin in solution, respectively) are independent of the cohesin under investigation and unaffected of whether or not Xyn-DocS is present during measurement. All cohesins under investigation show their expected and previously known unfolding force distributions without Xyn-DocS in solution<sup>4</sup> (middle column, orange). As expected, Coh3 and Coh5 can be fitted using the Bell-Evans model to obtain their most probable unfolding forces (dashed lines). The rupture force histograms of Coh1 and T107S were smoothed using kernel density estimations (KDE) to obtain most probable unfolding forces from them (dotted lines). The unfolding force distributions, and thus the Bell-Evans fits and KDEs, of all cohesins remain virtually unchanged upon the addition of Xyn-DocS (middle column, red, solid lines). Forces F on the left and right side of the middle column represent the most probable rupture forces of the cohesins without and with Xyn-DocS in solution, respectively, and their relative change in percent. Numbers N on the left and on the right side of the middle column represent the number of cohesin unfolding force data points in the histograms without and with dockerin in solution, respectively. We found that changing the bin size within a range of 10pN to 50pN did not notably alter the fits of the Bell-Evans model to the cohesin unfolding force histograms or the obtained most probable unfolding forces. We therefore chose the same bin size for all constructs within a column of histograms (25pN for all ddFLN4 and cohesin unfolding events, and 50pN for all CohIII-XDdocIII unbinding events), to provide good comparability by eye. All data were recorded using a single cantilever with a spring constant of 114pN/nm at a retraction speed of 1600 nm/s in a 24 hour automated SMFS experiment.

(a)		ddFLN4 unfolding forces F [pN]						(d)	
		most probable				mean		Number of curves	
		no doc	doc	$\Delta F$	no doc	doc	$\Delta F$	no doc	doc
Cohesin 1		78	78	0	80	77	-3	189	660
Cohesin 3		76	79	3	80	78	-2	197	783
Cohesin 5		75	76	1	80	78	-2	111	509
Cohesin T107S		73	79	5	80	78	-2	201	346
<b>average</b>		<b>76</b>	<b>78</b>	<b>2</b>	<b>80</b>	<b>78</b>	<b>-2</b>	<b>698</b>	<b>2298</b>

(b)		Cohesin unfolding forces F [pN]						(e)	
		most probable				mean		lever spring constant: 114pN/nm	
		no doc	doc	$\Delta F$	no doc	doc	$\Delta F$		
Cohesin 1		119	115	-4	151	143	-8	time w/o doc: ~6hrs	
Cohesin 3		296	296	1	277	277	0	time w/ doc: ~18hrs	
Cohesin 5		504	493	-11	468	465	-3	total exp time: ~24hrs	
Cohesin T107S*		105	277	109	273	3	-4		
<b>average</b>		<b>260</b>	<b>257</b>	<b>-3</b>	<b>280</b>	<b>278</b>	<b>-1</b>		

(c)		CohIII:XDdocIII unbinding forces F [pN]					
		most probable				mean	
		no doc	doc	$\Delta F$	no doc	doc	$\Delta F$
Cohesin 1		632	645	13	604	604	1
Cohesin 3		645	643	-2	605	607	2
Cohesin 5		648	650	2	624	615	-10
Cohesin T107S		637	646	9	602	606	4
<b>average</b>		<b>640</b>	<b>646</b>	<b>5</b>	<b>609</b>	<b>608</b>	<b>-1</b>

Table S3: Summary of all unfolding force data from negative control experiment. (a)-(c) Most probable and mean rupture for the ddFLN4 fingerprint, all cohesins of interest and the CohIII-XDdocIII pulling handle. None of the molecular constructs show a notable change in most probable or mean unfolding forces upon addition of the dockerin DocS. (d) Number of force curves of all molecular constructs. (e) General experiment information.

## Protein sequences

ybbR-tag - linker and additional residues - XylanaseT6 (B. stea) - linker - Doc124 (*A. cellulolyticus*):

MGTDSLEFIASKLALEVLFGQPLQHHPWTSASKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQML  
KRHFNSIVAENVMPISIQPEEGKFNEQADRIKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLL  
KRLETHIKTIVERYKDDIKYWDVNVVGGDDGKLRNSPWYQIAGIDYKVAFAAARKYGGDNIKLYMNDYNTEVEPKRTALYNL  
VKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYE  
KLSDKISNVTWFGIADNHTWLSRADVYYDANGNVVDPNAPYAKVEKGKGDAPFVFGPDYKVKPAYWAIIDHKVVPVAVTGD  
INGDGYFNSIDFGLMRVYLLSGSIPNYSAADVNGSDNSANSIDFGYMRQYLLGIITVFPNGGTQT

ybbR-tag - linker and additional residues - XylanaseT6 (B. stea) - linker - DocS (*C. thermocellum*):

MGTDSLEFIASKLALEVLFGQPLQHHPWTSASKNADSYAKKPHISALNAPQLDQRYKNEFTIGAAVEPYQLQNEKDVQML  
KRHFNSIVAENVMPISIQPEEGKFNEQADRIKFAKANGMDIRFHTLVWHSQVPQWFFLDKEGKPMVNECDPVKREQNKQLLL  
KRLETHIKTIVERYKDDIKYWDVNVVGGDDGKLRNSPWYQIAGIDYKVAFAAARKYGGDNIKLYMNDYNTEVEPKRTALYNL  
VKQLKEEGVPIDGIGHQSHIQIGWPSEAEIEKTINMFAALGLDNQITELDVSMYGWPPRAYPTYDAIPKQKFLDQAARYDRLFKLYE  
KLSDKISNVTWFGIADNHTWLSRADVYYDANGNVVDPNAPYAKVEKGKGDAPFVFGPDYKVKPAYWAIIDHKVVPVAVTGD  
KLYGDVNDGKVNSTDAVALKRYVLRSGIGSGSGSGSGSSINTDNADLNEDGRVNSTDLGILKRYILKEIDITLPYKN

ybbR-tag - linker and additional residues - Cohesin - linker - XDoc3:

MGTDSLEFIASKLALEVLFGQPLQHHPWTSAS

Cohesin 1

TGFTVNVDSVNGVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEADGKLVFLDYTMSTG  
YISTNGVFANVTFKVLNSAPITVIGITGATFGDKNLGNISATINAGSINGG

Cohesin 2

TGFTVNVDSVNGVGEQIVVPVSFANVPSNGISTADMTITYDSSKLEYVSGDAGSIVTNPTVNFNGINKETDGKLVFLDYTMSTG  
YISTNGVFAKVTFKVLNAGSSVIGITGATFGDKNLGVSATINAGSINGG

Cohesin 3

TGFTVNVDSVNGVGEQIVVPVSFANIPANGISTADMTITYDSSKLEYVSGVPGSIVTNPDVNFNGINKETDGKLVFLDYTMSTGYI  
STSGVFTKVTFKVLSGGSTVIGITGATFGDKNLGNVSATINAGSINGG

Cohesin 4

NAMAVAVGAVQGGVGETVTVPTMTKVPPTGVSTADFTVYDANKLEYVSGAAGSIVTNPDVNFNGINKEADGKIKVFLDYTM  
ATEYISKDGVFANLTFKIKSTAAAGTTAAVGIAGTATFGDSALKPITAVITDGKVEII

Cohesin 5

KAMKVVIANVSGNAGSEVVVPVSIEGVSANGVSAADFTITYDANKLEYVSGAAGSIVKNPDVNFNGINKEADGKLVFLDYTMAT  
EYISADGIFANLTFKIKSTAVNGDVA AISKSGTATFGDKNLGPISAVIKDGSVTVG

Cohesin 6

TGFNLSIDTVEGNPSSVVVVKLSGISKNGISTADFTVYDANKLEYISGDAGSIVTNPGVNFNGINKESDGKLVFLDYTMSTGYI  
STDGVFANLNFNIKSSAIGSKAEVSISGTPFGDSTLTPVAVKVTNGAVNVV

Cohesin 7

NAFKVSIDTVKAAATGTQVVVPVSFVNVPATGISTTDMTITYDANKLEYVSGDAGSIVTNPGVNFNGINKEADGKLVFLDYTMIT  
QYISEDGVFANVTFKVIKSTAVNGDVA AISKSGTATFGDSSLSPVTASVNVGGVNG

Cohesin 1 A105G P106G T107S "GGs"

TGFTVNVDSVNGVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEADGKLVFLDYTMSTG  
YISTNGVFANVTFKVLNSGGSTVIGITGATFGDKNLGNISATINAGSINGG

Cohesin 1 T107S

TGFTVNVDSVNGVGEQIVVPVSFANVPSNGVSTADMTITYDSSKLEYVSGAAGSIVTNPTVNFNGINKEADGKLVFLDYTMSTG  
YISTNGVFANVTFKVLNSAPITVIGITGATFGDKNLGNISATINAGSINGG

VVPNTVTSAVKTYVEIESVDGFYFNTEKFDTAQIKKAVLHTVYNEGVTGDDGVAVVLREYESEPVDTAELTFGDATPANTYK  
AVENKFDYEIPVYYNNATLKDAEGNDATVTVYIGLKGDTDLNNIVDGRDATATLTYYAATSDGKDATTVALSPSTLVGGNPESV  
YDDFSAFLSDVKVDAGKELTRFAKKAERLIDGRDASSILTFYTKSSVDQYKDMAANEPNKLWDIVTGDAEEE

Coh3 - linker - ddFLN4 - linker and additional residues - ybbR

MGTALDRGMTYDLDPKDGSSAATKPVLEVTKKVFDTAADAAGQTVTVEFKVSGAEGKYATTGYHIYWDERLEVVAKTGAY  
AKKGALEDSSLAKAENNGVVFVSGADDDFGADGVMWTVELKVPADAKAGDVYPIDVAYQWDPSKGLFTDNKDSAQ GK  
LMQAYFFFTQGIKSSSNPSTDEYLVKANATYADGYIAIKAGEPGSVVPSTGSADPEKSYAEGPGLDGGESFQPSKFKIHAVDPDGVH  
RTDGGDGFVVIEGPAPVDPVMVDNGDGTVDVEFEPKEAGDYVINLTLGDGNVNGFPKTVTVPAPGSELKLPRSRHHHHHHS  
LEVLFGQPDSEFIASKLA



### 3.3 PUBLICATION P3: From Genes to Protein Mechanics on a Chip

This article presents a parallelized, lab-on-a-chip-based sample preparation method, which allows high-throughput AFM-based SMFS data acquisition. It utilizes cell-free *in vitro* transcription and translation (IVTT) and site-specific, covalent protein pulldown in a microfluidic chip to prepare a protein microarray sample on a glass slide. The PDMS (Polydimethylsiloxane)-based microfluidics chip is based on an intricate dual-layer design, which allows for precise fluid control *via* pneumatically controlled valves. Each chip contains 640 double chambers within approximately 1 cm<sup>2</sup> for the expression and immobilization of theoretically hundreds of individual protein sample spots. This enables a new level of high-throughput mechanical phenotyping of libraries of protein domains.

# From Genes to Protein Mechanics on a Chip

Marcus Otten\*, Wolfgang Ott\*, Markus A. Jobst\*,  
Lukas F. Milles, Tobias Verdorfer, Diana A. Pippig,  
Michael A. Nash and Hermann E. Gaub

*\*these authors contributed equally to this publication*

*published in*

Nature Methods, 11(11), 1127-1130, (2014)

*Reprinted from [115], with permission from Nature Publishing Group  
Copyright 2014, Nature Publishing Group*

## From genes to protein mechanics on a chip

Marcus Otten<sup>1,2,4</sup>, Wolfgang Ott<sup>1,2,4</sup>, Markus A Jobst<sup>1,2,4</sup>, Lukas F Milles<sup>1,2</sup>, Tobias Verdorfer<sup>1,2</sup>, Diana A Pippig<sup>1-3</sup>, Michael A Nash<sup>1,2</sup> & Hermann E Gaub<sup>1,2</sup>

**Single-molecule force spectroscopy enables mechanical testing of individual proteins, but low experimental throughput limits the ability to screen constructs in parallel. We describe a microfluidic platform for on-chip expression, covalent surface attachment and measurement of single-molecule protein mechanical properties. A dockerin tag on each protein molecule allowed us to perform thousands of pulling cycles using a single cohesin-modified cantilever. The ability to synthesize and mechanically probe protein libraries enables high-throughput mechanical phenotyping.**

Mechanical forces play a pivotal role in biological systems by performing tasks such as guiding cell adhesion<sup>1</sup>, inducing gene expression patterns<sup>2</sup> and directing stem cell differentiation<sup>3</sup>. At the molecular level, mechanosensitive proteins act as sensors and transducers, communicating the presence and direction of applied forces to downstream signaling cascades. Conformational changes in response to mechanical forces<sup>4</sup> and energetic barriers along unfolding pathways can be probed by single-molecule force spectroscopy (SMFS) techniques<sup>4</sup>. Such techniques, including optical tweezers, magnetic tweezers and atomic force microscopy (AFM), have been used to interrogate high-affinity receptor-ligand binding<sup>5</sup>, measure unfolding and refolding dynamics of individual protein domains<sup>6-8</sup>, observe base-pair stepping of RNA polymerases<sup>9</sup> and identify DNA stretching and twisting moduli<sup>10</sup>.

Despite these successes, SMFS experiments have been limited by low throughput. Experimental data sets typically contain a majority of unusable force-distance traces owing to the measurement of multiple molecular interactions in parallel or a lack of specific interactions. Typical yields of interpretable single-molecule interaction traces in SMFS experiments vary between 1% and 25%. The incapacity of SMFS to quickly screen libraries of molecular variants has hindered progress toward understanding sequence-structure-function relationships at the single-molecule level. In particular, the need to prepare each protein sample and cantilever separately increases experimental workload and gives rise to calibration uncertainties. Therefore, methods to

interrogate the mechanical behavior of different proteins in a parallel and streamlined format with the same cantilever would offer distinct advantages. Such a screening approach could characterize single-molecule properties such as unfolding forces, interdomain mechanical signatures and mechanically activated catch-bond behavior<sup>1</sup>. Screening of these properties could find applications in biotechnology and human health studies in which mechanical dysregulation or misfolding is suspected to play a role in pathology<sup>11</sup>.

Here we developed a platform for parallel characterization of individual protein mechanics in a single experiment (Fig. 1). Microspotted gene arrays were used to synthesize fusion proteins *in situ* using cell-free gene expression. Proteins were covalently immobilized inside multilayer microfluidic circuits. A single cantilever was then positioned above the protein array and used to probe the mechanical response of each individual protein via a common C-terminal dockerin (Doc) fusion tag. Genes of interest were chosen such that each gene product exhibited an identifiable unfolding pattern when loaded from the N to the C terminus. Each target protein was expressed with an N-terminal 11-amino-acid ybbR tag, which was used to covalently and site-specifically link the protein to the surface via Sfp synthase-catalyzed reaction with coenzyme A (CoA)<sup>12</sup>. At the C terminus the proteins contained a 75-amino-acid cellulosomal Doc from *Clostridium thermocellum*<sup>13</sup> as a specific handle targeted by the cohesin (Coh)-modified cantilever.

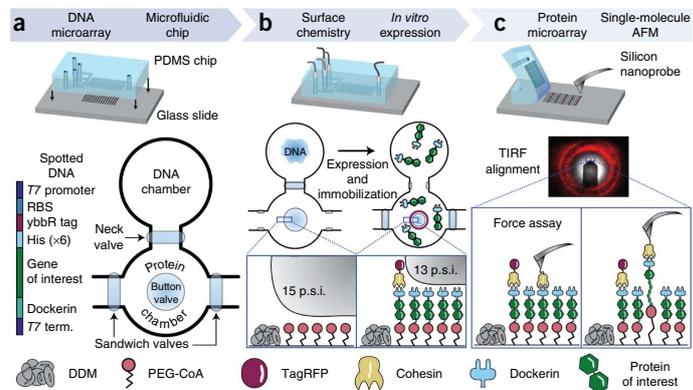
The gene microarray was aligned and reversibly bonded to a microfluidic chip known as MITOMI (mechanically induced trapping of molecular interactions). The chip has been used in the past for screening transcription factors<sup>14,15</sup> and mapping interaction networks<sup>16</sup>. More recently, our group employed MITOMI chips for molecular force assays<sup>17</sup>. In this work, MITOMI chips featured 640 dumbbell-shaped unit cells in a flow layer and 2,004 micromechanical valves in a control layer. Each unit cell was equipped with pneumatic 'neck', 'sandwich' and 'button' valves (Fig. 1a) according to design principles of soft lithography<sup>18</sup>. Each neck valve protected the microspotted DNA in the back chamber from exposure to other reagents during surface patterning in the front chamber. The sandwich valves prevented chamber-to-chamber cross contamination, ensuring that only a single protein variant was present in each sample spot. For surface chemistry in the front chamber, the button valves were actuated to shield the sample spots, allowing *n*-dodecyl  $\beta$ -D-maltoside passivation in the surrounding area. Releasing the button valves allowed subsequent functionalization with CoA-poly(ethylene glycol) (CoA-PEG) in the sample area under the buttons serving as the protein immobilization site. We expressed the genes by

<sup>1</sup>Lehrstuhl für Angewandte Physik, Ludwig-Maximilians-Universität, Munich, Germany. <sup>2</sup>Center for Nanoscience (CeNS), Ludwig-Maximilians-Universität, Munich, Germany. <sup>3</sup>Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität, Munich, Germany. <sup>4</sup>These authors contributed equally to this work. Correspondence should be addressed to M.A.N. (michael.nash@lmu.de).

RECEIVED 3 MARCH; ACCEPTED 29 JULY; PUBLISHED ONLINE 7 SEPTEMBER 2014; DOI:10.1038/NMETH.3099

## BRIEF COMMUNICATIONS

**Figure 1** | Method workflow. (a) A gene array was spotted onto a glass slide. Genes were designed with a common set of flanking sequences, including a 77 promoter region, ybbR tag, dockerin tag and 77 terminator (term.). The multilayer microfluidic chip featuring 640 unit cells was aligned to the DNA microarray and bonded to the glass slide. Each unit cell comprised a DNA chamber, a protein chamber, and superseding elastomeric control valves actuated by pneumatic pressure. PDMS, poly(dimethylsiloxane). (b) Control valves were used for spatially selective surface modification of each protein chamber with poly(ethylene glycol)-coenzyme A (PEG-CoA) and for fluidic isolation of each chamber before *in vitro* expression of the microspotted DNA. Fluorescence labeling with TagRFP-cohesin was achieved by partial button-valve pressurization, leaving only an outer concentric ring of immobilized gene products exposed to the labeling solution. DDM, *n*-dodecyl  $\beta$ -*D*-maltoside. (c) After removal of the microfluidic device, the resulting well-defined, covalently attached protein microarray was accessed from above with a cohesin-functionalized atomic force microscope (AFM) cantilever. Single-molecule unfolding traces of each of the protein constructs were thus acquired sequentially at each corresponding array address with a single cantilever in a single experiment.



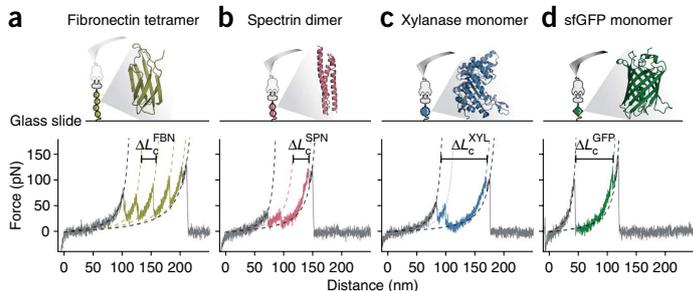
incubating an *in vitro* transcription and translation cell extract at 37 °C with the spotted DNA in the back chamber. The synthesized proteins then diffused to the front chamber, where they were covalently linked to the surface via an Sfp-catalyzed reaction of surface-bound CoA with solution-phase N-terminal ybbR peptide tags (Fig. 1b). Partial pressurization of the button valve<sup>19</sup> was used for tagging an outer concentric portion of the sample area with a fluorescently (TagRFP) tagged Coh that specifically bound to the C-terminal Doc tag of each target protein, thereby confirming successful protein synthesis and surface immobilization (Supplementary Fig. 1). Finally, the microfluidic device was removed from the glass slide to provide access to the protein array from above. Using this approach, we generated microarrays of site-specifically and covalently immobilized proteins for subsequent SMFS experiments, starting from a conventional gene array.

An inverted three-channel total-internal-reflection fluorescence/atomic force microscope (TIRF-AFM)<sup>20</sup> was used to position the cantilever in the center of the fluorescent rings in the protein array and perform SMFS measurements (Fig. 1c). The Coh-modified cantilever was used to probe the surface for expressed target proteins containing the C-terminal Doc tag. Upon surface contact of the cantilever, formation of a Coh-Doc

complex allowed measurement of target-protein unfolding in a well-controlled pulling geometry (N to C terminus). We retracted the probe at constant velocity and recorded force-extension traces that characterized the unfolding fingerprint of the target protein. This approach-retract process could be repeated many times at each array address to characterize each expression construct.

Several unique features of the C-terminal Doc tag make it particularly suitable as a protein handle for SMFS. Its small size of 8 kDa does not notably add to the molecular weight of the gene products, which is advantageous for cell-free expression. Additionally, Doc exhibits a specific and high-affinity interaction with Coh domains from the *C. thermocellum* scaffold protein CipA. Coh was used both for fluorescence detection of the expression constructs and for modification of the cantilever. On the basis of our prior work, the Coh-Doc interaction is characterized to be high affinity, with a dissociation constant  $K_d$  in the low nanomolar range and rupture forces >125 pN at a loading rate of 10 nN/s (ref. 21). Our prior work also indicated that upon forced dissociation, Doc exhibited a characteristic double sawtooth rupture peak with a contour length increment of 8 nm separating the two peaks. We used this two-pronged double rupture event at the end of each force-extension trace as a positive indicator

**Figure 2** | Representative single-molecule force traces recorded in different protein spots on a single chip with a single cantilever. (a-d) Four proteins of interest, anchored between the coenzyme A (CoA)-functionalized surface and the cohesin-functionalized cantilever, were probed: fibronectin tetramer (a), spectrin dimer (b), xylanase monomer (c) and sfGFP monomer (d). The crystal structure and pulling configuration (top) are shown for each construct. Each single-molecule force-distance trace (bottom) shows the individual unfolding fingerprint of the respective protein of interest followed by a common, final double sawtooth peak (gray) that is characteristic of the cohesin-dockerin rupture. Experimental data were fitted with the worm-like chain model (dashed lines). Unfolding intermediates were also observed (fitted for only xylanase in c; dotted colored line).



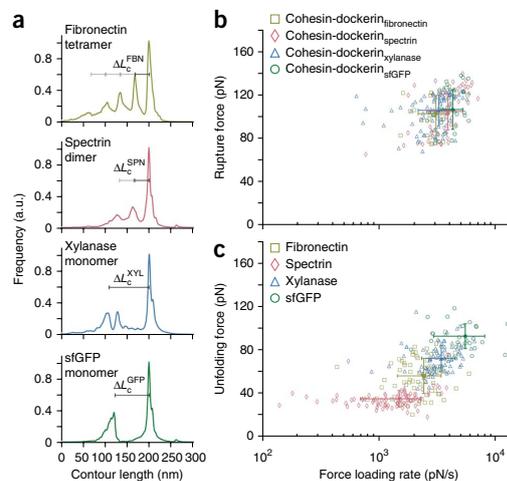
**Figure 3** | Unfolding and rupture statistics from multiple force traces. (a) Relative frequency of observing given contour lengths determined by transforming and aligning multiple force traces into contour length space via the worm-like chain model. Shown are diagrams for the fibronectin tetramer ( $n = 27$ ,  $\Delta L_c^{\text{FBN}} = 33$  nm), spectrin dimer ( $n = 50$ ,  $\Delta L_c^{\text{SPN}} = 34$  nm), xylanase monomer ( $n = 91$ ,  $\Delta L_c^{\text{XYL}} = 93$  nm) and sfGFP monomer ( $n = 25$ ,  $\Delta L_c^{\text{GFP}} = 79$  nm). (b) Rupture force versus loading rate of the final cohesin-dockerin dissociation event. (c) Unfolding force versus loading rate for each protein of interest. The populations in b and c were fitted with two-dimensional Gaussians. Respective means and s.d. are plotted in the corresponding colors as solid symbols and error bars. a.u., arbitrary units.

that the gene of interest was completely expressed through to the C terminus (Fig. 2). Furthermore, this double rupture peak indicated that the interaction with the Coh-modified cantilever was specific and that the pulling geometry was strictly controlled such that force was applied to the molecule of interest from the N to the C terminus.

To validate and demonstrate our approach, we expressed genes of interest comprising well-known fingerprint domains in the SMFS literature. We produced multimeric polyproteins including tetrameric human type-III fibronectin (FBN)<sup>22</sup> and dimeric chicken brain  $\alpha$ -spectrin (SPN)<sup>23</sup>. We also synthesized monomers of endo-1,4-xylanase T6 from *Geobacillus stearothermophilus* (XYL)<sup>21</sup>, superfolder GFP (GFP)<sup>24</sup> and twitchin kinase<sup>25</sup>. In all cases, the target proteins were fused to N-terminal ybbR and C-terminal Doc tags (Supplementary Figs. 2–6). Unfolding data for FBN, SPN, XYL and GFP were obtained using a single cantilever to probe a single microarray (Figs. 2 and 3). Twitchin kinase was found not to express in sufficient yield to provide reliable unfolding statistics.

We transformed force-extension data (Fig. 2) into contour length space<sup>26</sup> using the worm-like chain model and compared the measured contour length increments with the amino acid sequence lengths of each protein and literature values. The observed contour lengths and rupture forces were consistent with our expectations. FBN showed a fourfold-repeated sequence of rupture peaks at contour length increments of 32 nm ( $\Delta L_c^{\text{FBN}}$ ; Fig. 2a) frequently interrupted by an intermediate peak at 10–12 nm, both features characteristic of FBN<sup>22</sup>. SPN showed two regular sawtooth-like peaks with contour lengths of 33 nm ( $\Delta L_c^{\text{SPN}}$ ; Fig. 2b)<sup>23</sup>. XYL exhibited a decreasing multip peaked unfolding fingerprint with a contour length increment of 92 nm ( $\Delta L_c^{\text{XYL}}$ ; Fig. 2c), occasionally showing additional increments corresponding to unfolding of remaining XYL subdomains, a result consistent with the prior study and accounting for N-terminal immobilization of XYL<sup>21</sup>. GFP unfolding showed a contour length increment of 74 nm ( $\Delta L_c^{\text{GFP}}$ ; Fig. 2d)<sup>24</sup>. As each protein in the array contained the same C-terminal Doc tag, the final two rupture peaks in all force traces represented rupture of the Coh-Doc complex regardless of the protein of interest.

In our system, surface densities of expressed proteins were comparable to those obtained in conventional SMFS experiments. Uninterpretable and nonspecific interactions were excluded from the analysis (Supplementary Fig. 7). By collecting multiple unfolding traces, we assembled contour length diagrams for each protein of interest<sup>26,27</sup> (Fig. 3a) and confirmed the predicted contour length increments on the basis of the encoded amino acid sequences in each DNA spot. Coh-Doc rupture events for all protein constructs in the array clustered to the same population in the force-loading rate plot, independently of the preceding



rupture peaks from the protein of interest (Fig. 3b). The Coh-Doc ruptures agreed with previously reported values at similar loading rates<sup>21</sup>. The unfolding events of the proteins of interest produced distinct populations in the force-loading rate plots (Fig. 3c). The unfolding events depended on the internal structure and the unfolding pathway of the fingerprint domain when stretched between its N and C termini. SPN, for example, an elongated 3-helix bundle, was previously reported to exhibit a broader energy well ( $\Delta x = 1.7$  nm; ref. 23) and showed a flatter distribution of unfolding forces than that of the more compact globular FBN domain with a shorter, steeper potential ( $\Delta x = 0.4$  nm; ref. 22).

In summary, our flexible approach efficiently streamlines protein expression, purification and SMFS into a single integrated platform (Supplementary Discussion). The approach should be compatible with other *in vitro* expression systems including extracts derived from insects, rabbit reticulocytes and human cell lines, and it is capable of introducing post-translational modifications and non-natural amino acids, allowing, for example, the screening of site-directed mutants. Our method allows for synthesis of cytotoxic proteins or proteins with a tendency to form inclusion bodies during bulk expression. In addition to providing greatly improved throughput, our system has the advantage of measuring multiple constructs with one cantilever, thereby eliminating errors introduced when performing multiple calibrations on different samples with uncertainties of ~10% (ref. 28). Detecting subtle differences in mechanical stability with this high-throughput approach could therefore be used to perform mechanical phenotyping experiments on similarly stable families of mutant proteins. This workflow opens the door to large-scale screening studies of protein nanomechanical properties.

#### METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Addgene: pET28a-ybbR-HIS-sfGFP-DocI, 58708; pET28a-ybbR-HIS-CBM-CohI, 58709; pET28a-StrepII-TagRFP-CohI, 58710; pET28a-ybbR-HIS-Xyl-DocI, 58711;

## BRIEF COMMUNICATIONS

pET28a-ybbR-HIS-10FNIII(x4)-DocI, 58712; pET28a-ybbR-HIS-Spec(x2)-DocI, 58713.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

### ACKNOWLEDGMENTS

M.O. is grateful to the Elite Network of Bavaria (IDK-NBT) for a doctoral fellowship. M.A.N. acknowledges support from Society in Science—The Branco Weiss Fellowship administered by the ETH Zürich. The authors acknowledge support from the DFG Sonderforschungsbereich 1032 and the Excellence Cluster. The authors thank E. Bayer (Weizmann Institute) for starting genetic materials used for Doc and Coh modules.

### AUTHOR CONTRIBUTIONS

M.O., M.A.N. and H.E.G. designed the research; M.O., W.O., M.A.J. and T.V. performed experiments; D.A.P. helped with immobilization strategies; M.O., W.O., M.A.J., L.F.M. and M.A.N. performed data analysis; M.O., W.O., M.A.J., M.A.N. and H.E.G. cowrote the manuscript.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Thomas, W.E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E.V. *Cell* **109**, 913–923 (2002).
2. Li, C. & Xu, Q. *Cell. Signal.* **12**, 435–445 (2000).
3. Engler, A.J., Sen, S., Sweeney, H.L. & Discher, D.E. *Cell* **126**, 677–689 (2006).
4. Müller, D., Helenius, J., Alsteens, D. & Dufrêne, Y.F. *Nat. Chem. Biol.* **5**, 383–390 (2009).
5. Florin, E.-L., Moy, V.T. & Gaub, H.E. *Science* **264**, 415–417 (1994).
6. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. & Gaub, H. *Science* **276**, 1109–1112 (1997).
7. Fernandez, J.M. & Li, H. *Science* **303**, 1674–1678 (2004).
8. Oesterhelt, F. et al. *Science* **288**, 143–146 (2000).
9. Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R. & Block, S.M. *Nature* **438**, 460–465 (2005).
10. Bryant, Z. et al. *Nature* **424**, 338–341 (2003).
11. Linke, W.A. *Cardiovasc. Res.* **77**, 637–648 (2008).
12. Yin, J. et al. *Proc. Natl. Acad. Sci. USA* **102**, 15815–15820 (2005).
13. Bayer, E.A., Belaich, J.-P., Shoham, Y. & Lamed, R. *Annu. Rev. Microbiol.* **58**, 521–554 (2004).
14. Maerkl, S.J. & Quake, S.R. *Science* **315**, 233–237 (2007).
15. Rockel, S., Geertz, M., Hens, K., Deplancke, B. & Maerkl, S.J. *Nucleic Acids Res.* **41**, e52 (2013).
16. Gerber, D., Maerkl, S.J. & Quake, S.R. *Nat. Methods* **6**, 71–74 (2009).
17. Otten, M., Wolf, P. & Gaub, H.E. *Lab Chip* **13**, 4198–4204 (2013).
18. Thorsen, T., Maerkl, S.J. & Quake, S.R. *Science* **298**, 580–584 (2002).
19. Garcia-Cordero, J.L. & Maerkl, S.J. *Chem. Commun.* **49**, 1264–1266 (2013).
20. Gump, H., Stahl, S.W., Strackharn, M., Puchner, E.M. & Gaub, H.E. *Rev. Sci. Instrum.* **80**, 063704 (2009).
21. Stahl, S.W. et al. *Proc. Natl. Acad. Sci. USA* **109**, 20431–20436 (2012).
22. Li, L., Huang, H.H.-L., Badilla, C.L. & Fernandez, J.M. *J. Mol. Biol.* **345**, 817–826 (2005).
23. Rief, M., Pascual, J., Saraste, M. & Gaub, H.E. *J. Mol. Biol.* **286**, 553–561 (1999).
24. Dietz, H. & Rief, M. *Proc. Natl. Acad. Sci. USA* **103**, 1244–1247 (2006).
25. Greene, D.N. et al. *Biophys. J.* **95**, 1360–1370 (2008).
26. Puchner, E.M., Franzen, G., Gautel, M. & Gaub, H.E. *Biophys. J.* **95**, 426–434 (2008).
27. Jobst, M.A., Schoeler, C., Malinowska, K. & Nash, M.A. *J. Vis. Exp.* **82**, e50950 (2013).
28. Gibson, C.T., Smith, D.A. & Roberts, C.J. *Nanotechnology* **16**, 234–238 (2005).





## ONLINE METHODS

**Chip fabrication.** Ready-to-use wafers for flow and control layers of the 640-chamber MITOMI design were obtained from Stanford Microfluidics Foundry (design name DTPAd)<sup>14</sup>. The flow wafer features 15- $\mu\text{m}$ -high features, rounded by photoresist reflow, whereas the control wafer features a rectangular cross-section.

Microfluidic chips were cast in poly(dimethylsiloxane) (PDMS) from these wafers. For the control layer, Sylgard 184 (Dow Corning) base and curing agent were mixed at a ratio of 5:1 by weight, poured onto the wafer, degassed and partially cured for 20 min at 80 °C. For the flow-layer wafer, a 20:1 base-to-curing agent mixture of Sylgard 184 was spin-coated for 75 s at 1,600 r.p.m. and partially cured for 30 min at 80 °C. The control layer chips were cut out, inlet holes were punched and the chips were aligned onto the spin-coated PDMS on the flow-layer wafer. After the two-layer chips were baked for 90 min at 80 °C, they were cut and removed from the wafer, and inlet/outlet holes were punched. Microfluidic chips were stored for up to 6 weeks.

**Cloning.** For the construction of the fusion proteins, Gibson assembly<sup>29</sup> was used. A ratio of 0.07 pmol vector to 0.3 pmol of insert was used for the fusion reaction. The primer sequences are provided in **Supplementary Table 1**. A pET28a plasmid was linearized with primers 1 and 2. The dockerin type I-encoding gene was isolated from the xylanase-dockerin type I construct<sup>21</sup> with primers 3 and 4. Codon-optimized sequences were purchased from GeneArt/Invitrogen. The genes of interest were designed in such a way that they already contained sequences overlapping those of their neighboring partners (pET28a and dockerin type I). In the case of the spectrin, two domains were linked with a flexible glycine-serine ( $\times 6$ ) linker. For fibronectin, four type III domains were fused separated by glycine-serine ( $\times 6$ ) linkers. The expression vector in all cases was a pET28a plasmid with a modified multiple cloning site (sequence attached: plasmids are available at Addgene, **Supplementary Table 2**). After construction, clones were verified via sequencing and amplified in NEB 5- $\alpha$  *Escherichia coli* cells. Following plasmid preparation, samples were concentrated up to 500 ng/ $\mu\text{l}$  before microspotting.

**DNA microspotting.** A 24  $\times$  60-mm #1 thickness coverslip (Thermo Scientific) was silanized with 3-aminopropyltrimethyl-ethoxysilane (ABCR) following literature protocols<sup>30</sup>.

The DNA solution containing 1% (w/v) nuclease-free bovine serum albumin (Carl Roth) in nuclease-free water was microspotted under humid atmosphere onto the silanized coverslip using the GIX Microplotter II (Sonoplot) and a glass capillary with a 30- $\mu\text{m}$  tip diameter (World Precision Instruments) according to the manufacturer's instructions in a rectangular 40  $\times$  16 pattern with 320- $\mu\text{m}$  column pitch and 678- $\mu\text{m}$  row pitch. Alignment of the DNA array and the microfluidic chip was done manually using a stereomicroscope. Bonding between the glass cover slip and microfluidic device was achieved by thermal bonding for 5 h at 80 °C on a hot plate.

**Protein synthesis on-chip.** The microfluidic device was operated at a pressure of 4 p.s.i. in the flow layer and 15 p.s.i. in the control layer. Operation started with the button and neck valves actuated for surface passivation. The flow layer was passivated by flushing through standard buffer (25 mM Tris, 75 mM NaCl,

1 mM CaCl<sub>2</sub>, pH 7.2) for 5 min and 2% *n*-dodecyl  $\beta$ -D-maltoside (Thermo Scientific) in nuclease-free H<sub>2</sub>O for 30 min (ref. 31). Next the button valve was opened, and borate buffer (50 mM sodium borate, pH 8.5) was flushed through for 30 min to deprotonate aminosilane groups on the glass surface.

For maleimide/coenzyme A functionalization, a solution of 5 mM NHS-PEG-maleimide (MW = 513 Da, Thermo Scientific) in borate buffer was flushed through for 45 min. The device was then rinsed with nuclease-free H<sub>2</sub>O for 5 min, followed by 30 min of 20 mM coenzyme A (Merck) in coupling buffer (50 mM sodium phosphate, pH 7.2, 50 mM NaCl, 10 mM EDTA). The button valve was then actuated to protect the functionalized surface area followed by 5 min of rinsing with standard buffer.

S30 T7 HY (Promega) *in vitro* transcription and translation mix, supplemented with 1  $\mu\text{L}$  T7 polymerase (Promega) and 0.5  $\mu\text{L}$  RNase inhibitor (Invitrogen), was then flushed into the chip, filling the DNA chambers (neck valve open).

The neck valve was then closed, and the channels were filled with 4'-phosphopantetheinyl transferase (Sfp synthase) in Sfp buffer (50 mM HEPES, 10 mM MgCl<sub>2</sub>). The chip was then incubated at 37 °C on a hot plate. After 1 h of incubation, the neck and the button valves were opened to allow Sfp synthase-catalyzed linkage of expressed protein to the coenzyme A-functionalized area below the button. At the same time the sandwich valves were actuated to avoid chamber-to-chamber cross-contamination. After another 1.5 h of incubation, the neck and button valves were closed, the sandwich valves were opened and the chip was rinsed with standard buffer for 20 min.

To verify successful protein expression and immobilization on the functionalized surface area, a fluorescent detection construct (TagRFP-cohesin type I (2  $\mu\text{g}/\text{ml}$ ) in standard buffer) was flushed through the device for 10 min with the button valve actuated. The sandwich valves were then actuated, and the button valve partially released by decreasing the pressure to 11 p.s.i. After 20 min of incubation at room temperature, the sandwich valves were opened, and the chip flushed with standard buffer for 20 min. Fluorescence images of all chambers were recorded on an inverted microscope with a 10 $\times$  objective (Carl Zeiss), featuring an electron-multiplying charge-coupled device (EMCCD) camera (Andor). Prior to force spectroscopy experiments, the chip was stored in buffer at 4 °C.

Directly before measurement, the PDMS chip was peeled off from the glass substrate under buffer, revealing the microarray while avoiding drying of the functionalized surface. The array surface was then rinsed several times with buffer. We did not encounter any problems with cross-contamination between chambers.

**Cantilever functionalization.** A silicon-nitride cantilever bearing a silicon tip with a tip radius of  $\sim 8$  nm (Biolever mini, Olympus) was silanized with ABCR as described previously<sup>30</sup>. Protein functionalization was performed in a similar way as reported previously<sup>27,31</sup>. Briefly, a 50  $\mu\text{M}$  solution of CBM A2C-cohesin from *C. thermocellum* in standard buffer was incubated with 1:2 (v/v) TCEP beads (Tris (2-carboxyethyl) phosphine disulfide reducing gel, Thermo Scientific), previously washed with standard buffer, for 2.5 h. The cantilever was submerged in borate buffer for 45 min to deprotonate primary amine groups on the silanized surface and then incubated with 20 mM NHS-PEG-maleimide (MW = 5 kDa, Rapp Polymere) in borate buffer for 60 min.



The cantilever was rinsed sequentially in three beakers of deionized H<sub>2</sub>O. TCEP beads were separated from the protein solution by centrifugation at 1,000g for 1 min. Next the cantilever was incubated for 60 min with reduced protein solution, which was diluted to a concentration of 1 mg/mL with standard buffer. Finally the cantilever was rinsed sequentially in three beakers of standard buffer and stored submerged in standard buffer in humid atmosphere at 4 °C for up to 24 h before use.

**Force spectroscopy.** A custom-built TIRF (total internal reflection fluorescence)-AFM (atomic force microscope) hybrid<sup>20,30</sup> was used to conduct the force spectroscopy measurements. The TIRF microscope was used to image fluorophores in up to three different color channels simultaneously using an iChrome MLE-S four-color laser (Toptica Photonics), an Optosplit III triple emission image splitter (Cairn Research) and a Xion3 EMCCD camera (Andor). A long-range stick-slip *xy* piezo nanopositioning system (ANC350, Attocube Systems) allowed access to the whole microchip array as well as fine spatial sampling of different surface molecules on the nanometer scale within each protein spot. Cantilever actuation in the *z* direction was performed by a LISA piezo-actuator (Physik Instrumente) driven by an MFP3D AFM controller (Asylum Research).

The following force spectroscopy protocol was performed repeatedly in each functionalized protein target area. The cantilever approach velocity was 3,000 nm/s, dwell time at the surface was 10 ms and retract velocity was 800 nm/s. Data were recorded with 6,250-Hz sampling rate. The cantilever typically had a spring constant in the range of 100 pN/nm and a resonance frequency of 25 kHz in water. Accurate calibration of the system was performed by the nondestructive thermal method<sup>32,33</sup> using corrections to account for discrepancies from the original theory<sup>27,34</sup>.

**Data and statistical analysis.** The raw data were converted from photodiode voltages into force values in newtons, and the following standard corrections were applied. The zero force value for the unloaded cantilever in each curve was determined by averaging over 40-nm extension after the final complex rupture and subtracting this value from each force value in the curve. The position of the surface was determined by finding the force value closest to 0 in a small neighborhood of the first non-negative force value in the force-extension trace. The *z* piezo position was corrected for the true tip-sample separation due to deflection of the lever as a function of the force for a Hookean spring.

A pattern-recognition software based on a package described previously<sup>26</sup> and adapted in-house chose the curves showing worm-like chain force responses of the stretched protein constructs. Example curves showing multiple, unspecific or no interactions are shown in **Supplementary Figure 7**, together with a single xylanase trace for comparison. The expected protein backbone contour length increments for each construct were detected in contour length space: the real part of the following numerically solved inverse worm-like chain (WLC) formula<sup>27</sup>

was used to transform force-extension data into force–contour length space for every measured force curve:

$$L(x) = \frac{x}{6u} \left( 3 + 4u + \frac{9 - 3u + 4u^2}{g(u)} + g(u) \right)$$

where

$$g(u) = \left( 27 - \frac{27}{2}u + 36u^2 - 8u^3 + \frac{3\sqrt{3}}{2} \sqrt{-u^2((4u-3)^3 - 108)} \right)^{\frac{1}{3}}$$

and

$$u = F \frac{L_p}{kT}$$

with *L* the contour length, *x* the extension, *F* the force, *L<sub>p</sub>* the persistence length, *k* Boltzmann's constant and *T* the temperature. Transformed data points were combined in a Gaussian kernel density estimate with a bandwidth of 1 nm and plotted with a resolution of 1 nm. In these resulting energy-barrier position diagrams, the contour length increments could easily be determined. The transformation was performed with the following parameters: persistence length *L<sub>p</sub>* = 0.4 nm, thermal energy *kT* = 4.1 pN nm. Force and distance thresholds were applied at 10 pN and 5 nm, respectively. The measurement data sets in each protein spot on the chip typically showed a yield of 0.5–5% specific interactions.

The force peaks corresponding to protein domain unfolding events, as well as those corresponding to final ruptures, were line fitted in force-time space to measure the loading rate of each individual event.

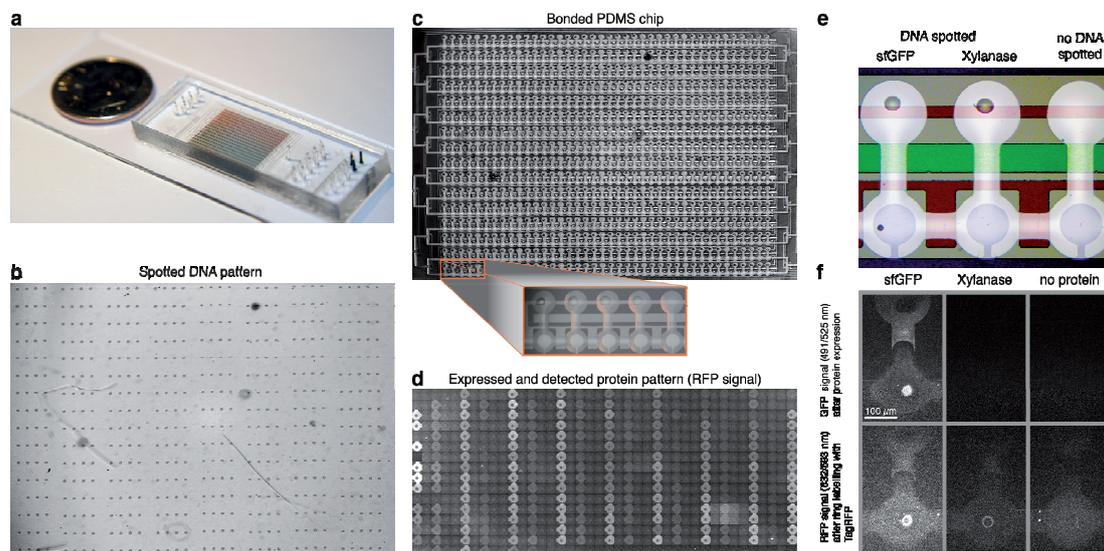
WLC fits for demonstrative purposes in **Figure 2** were done by using the following formula:

$$F(x) = \frac{kT}{L_p} \left( \frac{1}{4(1-x/L)^2} + \frac{x}{L} - \frac{1}{4} \right)$$

with *F* the force, *k* the Boltzmann's constant, *T* the temperature, *L<sub>p</sub>* the persistence length, *x* the extension and *L* the contour length.

Discrepancies between contour length increments in fitted single-molecule traces and aligned contour length diagrams are artifacts caused by the fixed persistence length in the contour length transformation, whereas the WLC fits to single force traces treat both contour length and persistence length of each stretch as free parameters. An overview of the yield of interpretable curves of all constructs is available in **Supplementary Table 3**.

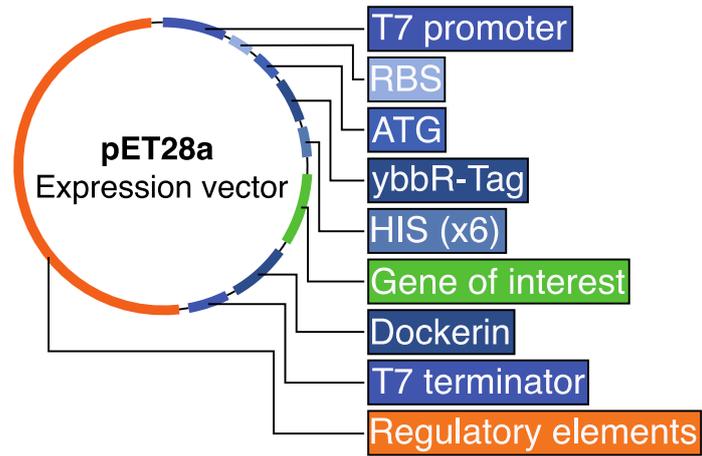
29. Gibson, D.G. *et al. Nat. Methods* **6**, 343–345 (2009).
30. Zimmermann, J.L., Nicolaus, T., Neuert, G. & Blank, K. *Nat. Protoc.* **5**, 975–985 (2010).
31. Huang, B., Wu, H., Kim, S. & Zare, R.N. *Lab Chip* **5**, 1005–1007 (2005).
32. Hutter, J.L. & Bechhoefer, J. *Rev. Sci. Instrum.* **64**, 1868 (1993).
33. Cook, S.M. *et al. Nanotechnology* **17**, 2135–2145 (2006).
34. Proksch, R., Schäffer, T.E., Cleveland, J.P., Callahan, R.C. & Viani, M.B. *Nanotechnology* **15**, 1344–1350 (2004).



### Supplementary Figure 1

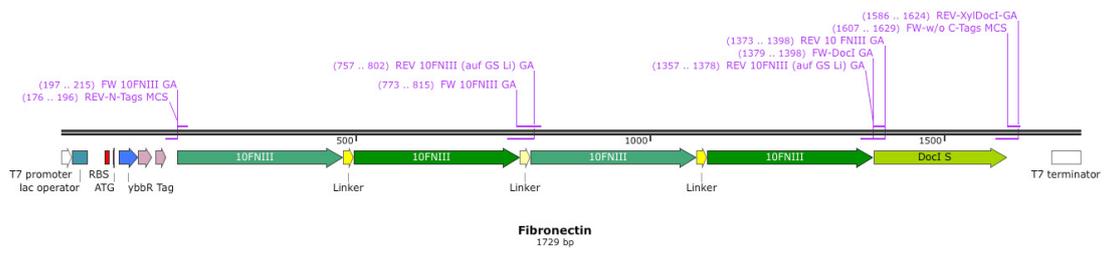
Microfluidic chip overview.

(a) Photograph of a microfluidic chip bonded to a glass slide with a US dime for scale. Control channels are filled with food dye for better visualization. (b) Pattern of a typical DNA array, consisting of repeats of rows with four different genes and one row with nothing spotted as negative control. (c) Photograph of a bonded PDMS chip onto the glass slide with DNA spots in the back chamber. The orange highlighted frame shows a zoom in of the bottom left corner. (d) Typical fluorescence collage assembled from 640 single fluorescence micrographs of each protein chamber on one single chip shows pattern of expressed protein (assembly not to scale). Fluorescence signal of TagRFP reveals expression levels and Dockerin specificity. Here, low passivation of the protein chamber facilitates visualization. (e) Three of 640 adjacent dumbbell-shaped chambers, one with sfGFP DNA spotted (left), one with Xylanase DNA (center) and one negative control without DNA (right). Control channels are visualized with food dye: neck valve (green), sandwich valve (red), and button valve (blue). (f) Fluorescence images showing GFP signal (top) from expressed and immobilized ybbR-sfGFP-Dockerin (left), ybbR-Xylanase-Dockerin (center) with negative control lacking the spotted DNA (right). The bottom row shows the signal from the TagRFP detection construct, which specifically bound to the Dockerin tag via the Cohesin domain.



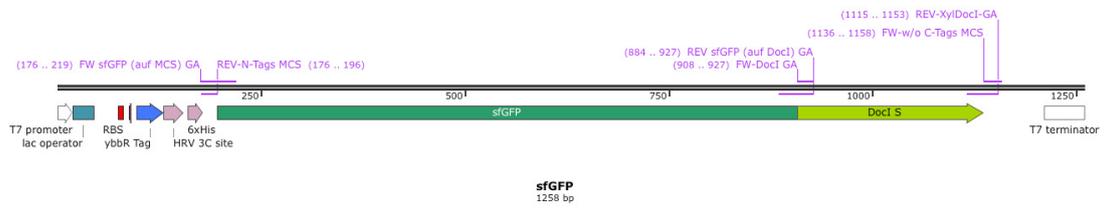
**Supplementary Figure 2**

Diagram of the expression vector pET28a with an individual gene of interest.



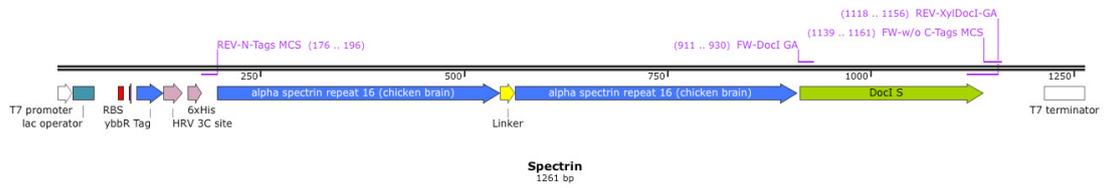
**Supplementary Figure 3**

Schematic of the fibronectin tetramer gene cassette.



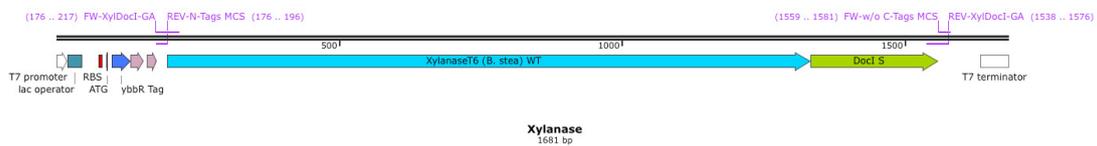
**Supplementary Figure 4**

Schematic of the sfGFP dimer gene cassette.



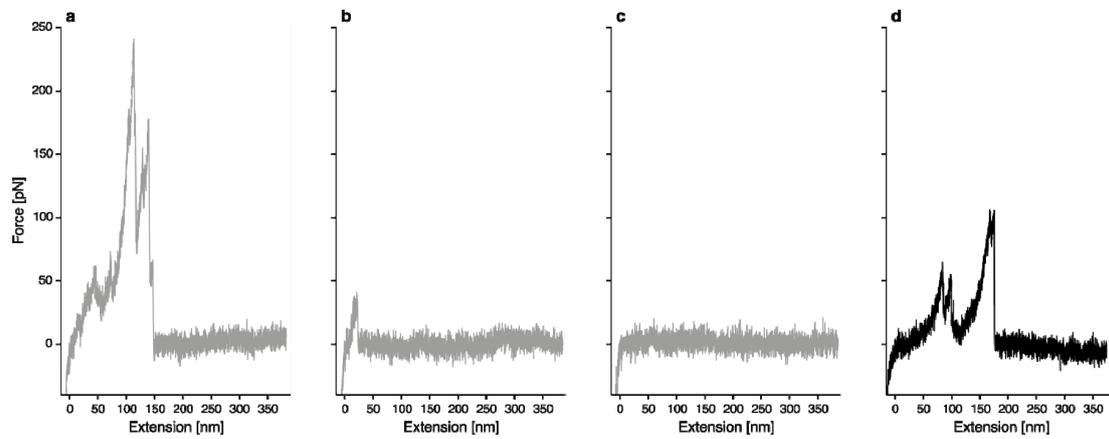
**Supplementary Figure 5**

Schematic of the spectrin dimer gene cassette.



**Supplementary Figure 6**

Schematic of the xylanase gene cassette.



### Supplementary Figure 7

Exemplary force traces

Example curves showing (a) uninterpretable interaction, (b) non-specific interaction of cantilever with surface, (c) no interaction, and (d) a specific Xylanase-Dockerin unfolding and unbinding trace. Curves similar to those shown in a-c were excluded from the analysis.

## Supplementary Discussion

Typically in SMFS experiments, rupture force – loading rate plots are used to characterize  $k_{\text{off}}$  and  $\Delta x$ , the unbinding (or unfolding) probability per time unit and the distance to the transition state along the reaction coordinate, respectively, providing direct information about the energy landscape governing protein folding<sup>1</sup>. SMFS experiments are also complemented by all-atom simulations of such systems *in silico*. Recently, it was shown that high speed SMFS experiments could be performed at speeds achievable in molecular dynamics simulations<sup>2</sup>, overcoming a long standing discrepancy between experiment and simulation.

In analyzing single-molecule unfolding curves (i.e., **Fig. 2**), we note that the spotted DNA at the measured array addresses correctly corresponded to the domain of interest encoded by the corresponding spotted DNA at that position. For example, the fibronectin tetramer was measured at array position (237), the spectrin dimer at position (239), the xylanase monomer at position (196), and the sfGFP monomer at position (238), corresponding to the correct genes deposited into the expression chambers at those array positions (**Fig. 2**). Typically 10–15 immobilization chambers per microarray were measured. Typically several thousand force curves were acquired giving rise to dozens of interpretable single-molecule interaction curves.

### Upper force limit

Here we extend the discussion regarding the upper force limit for the SMFS-MITOMI system. In all force-distance data traces, the last rupture events represent unbinding of the Coh-Doc complex, not unfolding of a domain. This rupture force of the Coh-Doc complex represents an upper limit in force for the entire construct, since the Doc is used as a handle sequence grabbed by the Coh-modified cantilever. The system we described can therefore interrogate domains with mechanical rupture forces that lie below that of Coh-Doc (~125 pN at 10 nN/s). If proteins with larger unfolding forces should be investigated, other Coh-Doc domains that show even higher complex rupture forces can be used. The Coh-Doc pair from *R. flavefaciens*, for example (PDB 4IU3) exhibits rupture forces over 600 pN at these loading rates (unpublished data). This could alternatively be used as a handle sequence to interrogate mechanically more stable domains of interest.

Computerized image analysis can be used to automate cantilever positioning above the fluorescent rings and subsequent acquisition of unfolding traces at each array address in combination with online force curve analysis to further increase throughput. Additionally, well-characterized reference proteins on the same chip may serve as calibration standards further minimizing uncertainty in absolute force values.

It is possible to operate the MITOMI device in a simplified way without the need for microspotting template DNA and chip alignment. This manual option should encourage the interested community to apply the suggested method to their single molecule force spectroscopy experiments. MITOMI enables the experimenter to prepare up to 16 different constructs in one column with 40 repeats each by flow-loading the DNA. Since

the valves are pressure sensitive it is also possible to operate these manually. This way it is possible to make use of the parallelized method without having the automation tools. Supplementary Materials & Methods

### DNA Sequences

**Supplementary Table 1.** Overview of primers

	Name	Sequence
1	FW-w/o C-Tags MCS	TAACCTCGAGTAAGATCCGGCTGC
2	REV-N-Tags MCS	GCTAGCACTAGTCCATGGGTG
3	FW-Docl GA	AAAGTGGTACCTGGTACTCC
4	REV-XylDocl-GA	CGGATCTTACTCGAGTTAGTTCCTGTACGGCAATGTATC
5	FW 10FNIII GA	CGCACCGGCTCTGGCTCTGGCTCTGTTAGTGATGTTCCGCGTG
6	REV 10 FNIII GA	GGAGTACCAGGTACCACTTTGGTGCG
7	REV 10FNIII (auf GS Li) GA	ACTAACAGAGCCAGAGCCAGAGCCGGTGCGATAATTGATTGAAATC
8	FW sfGFP (auf MCS) GA	CACCCATGGACTAGTGCTAGCAGCAAAGGTGAAGAACTGTTTAC
9	REV sfGFP (auf Docl) GA	GGAGTACCAGGTACCACTTTCTTATACAGCTCATCCATACCATG

**Supplementary Table 2.** Overview of DNA plasmids available at Addgene database

Addgene ID	Construct
58708	pET28a-ybbR-HIS-sfGFP-Docl
58709	pET28a-ybbR-HIS-CBM-CohI
58710	pET28a-StrepII-TagRFP-CohI
58711	pET28a-ybbR-HIS-Xyl-Docl
58712	pET28a-ybbR-HIS-10FNIII(x4)-Docl
58713	pET28a-ybbR-HIS-Spec(x2)-Docl

Multiple cloning site for the protein of interest:

*N terminal region*

*T7 promoter* | *lac operator* | *RBS* | *ATG* | *ybbr Tag* | *HRV 3C protease site* | *HIS Tag (x6)*

TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT  
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|GACTCTCTGGAATTCATCGCTTCTAA  
ACTGGCT|CTGGAAGTTCGTTCAGGGTCCG|CTGCAG|CACCACCACCACCAC|CCATGG  
ACTAGTGCTAGC

*C terminal region*

*Dockerin Type I* | *T7 terminator*

AAAGTGGTACCTGGTACTCCTTCTACTAAATTATACGGCGACGTCAATGATGACGGAAAAGTTAA  
CTCAACTGACGCTGTAGCATTGAAGAGATATGTTTTGAGATCAGGTATAAGCATCAACACTGACA  
ATGCCGATTTGAATGAAGACGGCAGAGTTAATTCAACTGACTTAGGAATTTTGAAGAGATATATT  
CTCAAAGAAATAGATACATTGCCGTACAAGAAC|TAA|CTCGAGTAAGATCCGGCTGCTAACAAA  
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGG  
CCTCTAAACGGGTCTTGAGGGGTTTTTT

## 10 FibronectinIII (4x):

**Glycin-Serin Linker (x6)**

GTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAGCCTGCTGATTAGCTG  
GGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAACCGGTGGTAATAGTC  
CGGTTCAAGAATTTACCGTTCGCGGTAGCAAAGCACCAGCAACCTAGCGGTCTGAAACCGGGT  
GTTGATTACACCATACCGTTTATGCCGTTACCGGTGCTGGTGATTCACCGGCAAGCAGCAAACC  
GATTAGCATTAACATATCGTACC**GGTAGCGGTAGTGGTAGC**GTTTCAGATGTGCCTCGCGACCTGG  
AAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCTGCGGTGACCGTGCGC  
TATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGAATTCACAGTTCCTGG  
TTCAAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACACAATCACAGTGTATG  
CAGTGACAGGTGCGGGTATAGTCCGGCAAGTTCAAAACCGATTCATCAATTAATCGCACCGGC  
**TCTGGCTCTGGCTCT**GTTAGTGATGTTCCGCGTGATCTGGAAGTTGTTGCAGCAACCCCGACCAG  
CCTGCTGATTAGCTGGGATGCACCGGCAGTTACCGTTCGTTATTATCGTATTACCTATGGTGAAA  
CCGGTGGTAATAGTCCGGTTCAGAATTTACCGTTCGCGGTAGCAAAGCACCAGCAACCTAGC  
GGTCTGAAACCGGGTGTGATTACACCATACCGTTTATGCCGTTACCGGTGCTGGTGATTACAC  
GGCAAGCAGCAAACCGATTAGCATTAACATATCGTACC**GGTAGCGGTAGTGGTAGC**GTTTCAGATG  
TGCCTCGCGACCTGGAAGTGGTGGCTGCCACACCGACCTCACTGCTGATCTCATGGGATGCCCT  
GCCGTGACCGTGCGCTATTATCGCATCACATATGGCGAGACAGGTGGCAATTCACCTGTGCAAGA  
ATTCACAGTTCCTGGTTCAAAAGTACCGCCACAATTTCTGGCCTGAAACCTGGCGTGGATTACA  
CAATCACAGTGTATGCAGTGACAGGTGCGGGTATAGTCCGGCAAGTTCAAAACCGATTCATC  
AAttaTCGCACC

**sfGFP:**

AGCAAAGGTGAAGAAGCTGTTTACCGGTGTTGTTCCGATTCTGGTTGAACTGGATGGTGTATGTTAA  
TGGCCACAAATTTTCAGTTCGTGGTGAAGGCGAAGGTGATGCAACCATTGGTAAACTGACCCTGA  
AATTTATCTGTACCACCGGCAAACTGCCGGTTCGGTGGCCGACCCTGGTTACCACCTGACCTAT  
GGTGTTCAGTGTTTTAGCCGTTATCCGGATCATATGAAACGCCACGATTTTTTCAAAGCGCAAT  
GCCGGAAGTTATGTTCAAGAACGTACCATCTCCTTTAAAGACGACGGTAAATACAAAACCGTG  
CCGTTGTTAAATTTGAAGGTGATACCCTGGTGAATCGCATTGAAGTGAAGGCACCGATTTTAAA  
GAGGATGGTAATATCCTGGGCCACAACTGGAATATAATTTCAATAGCCACAACGTATATCAC  
CGCAGACAAACAGAAAAATGGCATCAAAGCCAATTTTACCGTGCGCCATAATGTTGAAGATGGTA  
GCGTGCAGCTGGCAGATCATATCAGCAGAATACCCCGATTGGTGTATGGTCCGGTTCCTGCTGCCG  
GATAATCATTATCTGAGCACCCAGACCGTTCCTGAGCAAAGATCCGAATGAAAAACGTGATCATAT  
GGTGTGCATGAGTATGTTAATGCAGCAGGTATTACCCTGGTATGGATGAGCTGTATAAG

**alpha-Spectrin repeat 16 (chicken brain) (x2):**

**Glycin-Serine Linker (x6)**

CGTGTAAACTGAACGAATCTCACCGTCTGCACCAGTTCCTCCGTGACATGGACGACGAAGAATC  
TTGGATCAAAGAAAAAACTGCTGGTTCCTCTGAAGACTACGGTCGTGACCTGACCGGTGTTT  
AGAACCTGCGTAAAAAACACAAACGTCTGGAAGCTGAACCTGGCTGCTCACGAACCGGTATCCAG  
GGTGTTCCTGGACACCGGTAAAAAACTGTCTGACGACAACACCATCGGTAAAAGAAGAAATCCAGCA  
GCGTCTGGCTCAGTTCGTTGACCCTGGAAGAAGTGAACAGCTGGCTGCTGCTCGTGGTTCAGC  
GTCGGAAGAATCTCTGGAATACGGTAGCGGTAGCGGTTCACTGCTAAACTGAACGAATCTCAC  
CGTCTGCACCAGTTCCTCCGTGACATGGACGACGAAGAATCTTGGATCAAAGAAAAAACTGCT  
GGTTCCTCTGAAGACTACGGTCGTGACCTGACCGGTGTTTCAGAACCTGCGTAAAAAACACAAAC  
GTCGGAAGCTGAACCTGGCTGCTCACGAACCGGTATCCAGGGTGTCTGGACACCGGTAAAAAA  
CTGTCTGACGACAACACCATCGGTAAAAGAAGAAATCCAGCAGCGTCTGGCTCAGTTCGTTGACCA  
CTGGAAGAAGTGAACAGCTGGCTGCTGCTCGTGGTTCAGCGTCTGGAAGAATCTCTGGAATAT

**Xylanase:**

AAGAATGCAGATTCTATGCGAAAAAACCTCACATCAGCGCATTGAATGCCCCACAATTGGATCA  
ACGCTACAAAACGAGTTCACGATTGGTGCAGCAGTAGAACCTTATCAACTACAAAATGAAAAAG  
ACGTACAAATGCTAAAGCGCCACTTCAACAGCATTGTTGCCGAGAACGTAATGAAACCGATCAGC  
ATTC AACCTGAGGAAGGAAAAATCAATTTTGAACAAGCGGATCGAATGTTGAAGTTCGCTAAGGC  
AAATGGCATGGATATTCGCTTCCATACACTCGTTCGACAGCAAGTACCTCAATGGTTCCTTTC  
TTGACAAGGAAGGTAAGCCAATGGTTAATGAATGCGATCCAGTGAACAGTGAACAAAATAAACAA  
CTGCTGTTAAAACGACTTGAAACTCATATTAACAGATCGTTCGAGCGGTACAAAGATGACATTAA  
GTACTGGGACGTTGTAATGAGGTTGTGGGGGACGACGGAAAACTGCGCAACTCTCCATGGTATC  
AAATCGCCGGCATCGATTATATTAAGTGGCATTCCAAGCAGCTAGAAAATATGGCGGAGACAAC  
ATTAAGCTTTACATGAATGATTACAATACAGAACTCGAACCGAAGCGAACCGCTCTTACAATTT  
AGTCAAACTGAAAGAAGAGGGTGTCCGATCGACGGCATCGGCCATCAATCCCACATCCAAA  
TCGGCTGGCCTTCTGAAGCAGAAATCGAGAAAACGATTAACATGTTTCGCCGCTCTCGGTTTAGAC  
AACCAATCACTGAGCTTGATGTGAGCATGTACGGTTGGCCGCGCGCTTACCCGACGTATGA  
CGCCATTCAAAACAAAAGTTTTTGGATCAGGCAGCGCGCTATGATCGTTTGTCAAACGTATG  
AAAAGTTGAGCGATAAAATAGCAACGTACCTTCTGGGGCATCGCCGACAATCATACGTGGCTC  
GACAGCGTGCAGGATGTACTATGACGCCAACGGGAATGTTGTGGTTGACCCGAACGCTCCGTA  
CGCAAAAGTGGAAAAGGAAAAGATGCGCCGTTCTGTTTGGACCGGATTACAAAAGTCA  
AACCCGCATATTGGGCTATTATCGACCAC

## Detection construct RFP-Cohesin:

TagRFP-Cohesin:

*T7 promoter* | *lac operator* | *RBS* | *ATG* | *StrepII Tag* | *TagRFP* |  
*Linker* | *Cohesin* | *T7 terminator*

```
TAATACGACTCACTATAGG|GGAATTGTGAGCGGATAACAATTCC|CCTGTAGAAATAATTTTGT  
TTAACTTTAAG|AAGGA|GATATACAT|ATG|GGTACC|TGGTCTCACCCGCAGTTCGAAAAA|G  
TTTCTAAAGGTGAAGAACTGATCAAAGAAAACATGCACATGAAACTGTACATGGAAGGTACTGTT  
AACAAACCACCACTTCAAATGCACCTCTGAAGGTGAAGGTAACCGTACGAAGGTACTCAGACCAT  
GCGTATCAAAGTTGTTGAAGGTGGTCCGCTGCCGTTTCGCTTTTCGACATCCTGGCTACCTCTTTCA  
TGTACGGTTCTCGTACCTTCATCAACCACACCCAGGGTATCCCGGACTTCTTCAAACAGTCTTTC  
CCGGAAGGTTTCACCTGGGAACGTGTTACCACCTACGAAGACGGTGGTGTCTGACCGCTACCCA  
GGACACCTCTCTGCAAGACGGTTGCCTGATCTACAACGTTAAAAATCCGTGGTGTAACTTCCCCT  
CTAACGGTCCGGTTATGCAGAAAAAACCTGGGTTGGGAAGCTAACACCGAAATGCTGTACCCG  
GCTGACGGTGGTCTGGAAGGTCGTTCTGACATGGCTCTGAACTGGTGGTGGTGGTCACTGAT  
CTGCAACTTCAAACCACCTACCGTTCTAAAAAACCGGCTAAAACTGAAAATGCCGGGTGTTT  
ACTACGTTGACCACCGTCTGGAACGTATCAAAGAAGCTGACAAAGAACTACGTTGAACAGCAC  
GAAGTTGCTGTTGCTCGTTACTGCGACCTGCCGCTAAACTGGGTACAAACTGAAC|GGCAGTG  
TAGTACCATCAACACAGCCTGTAACAACACCACCTGCAACAACAAAACCACCTGCAACAACAATA  
CCGCCGTCAGATGATCCGAATGCA|GGATCCGACGGTGTGGTAGTAGAAATTGGCAAAGTTACGG  
GATCTGTTGGAAGTACAGTTGAAATACCTGTATATTTAGAGGAGTTCATCCAAAGGAATAGCA  
AATGCGACTTTGTGTTTCAGATATGATCCGAATGTATTGGAAATTATAGGGATAGATCCCGGAGA  
CATAATAGTTGACCCGAATCCTACCAAGAGCTTTGATACTGCAATATATCCTGACAGAAAGATAA  
TAGTATTCCTGTTTGGCGAAGACAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTT  
GCAAAAATAAGAGCAACTGTAAAATCAAGTGCTCCGGGTATATTAATTTTCGACGAAGTAGGTGG  
ATTTGCAGATAATGACCTGGTAGAACGAAAGGTATCATTTATAGACGGTGGTGTAACTGTTGGCA  
ATGCAACA|TAA|CTCGAGTAAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTG  
CTGCCACCGCTGAGCAATAA|CTAGCATAACCCCTTGGGGCCTCTAAACCGGCTTTGAGGGGTTT  
TTT
```

## Molecular weights of synthesized fusion proteins

ybbR-(Fibronectin)<sub>4</sub>-Dockerin Type I: 53 kDa  
ybbR-(Spectrin)<sub>2</sub>-Dockerin Type I: 40 kDa  
ybbR-Xylanase-Dockerin Type I: 56 kDa  
ybbR-sfGFP-Dockerin Type I: 39 kDa  
ybbR-Twitchin-Dockerin Type I: 52 kDa

**Supplementary Table 3.** Yield of interpretable curves

<b>Construct</b>	<b>Interpretable Curves</b>
GFP	25 out of 15258 = 0.16 %
Fibronectin	27 out of 26653 = 0.1 %
Xylanase	91 out of 5553 = 1.64 %
Spectrin	50 out of 10344 = 0.48%

## References

1. Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50–53 (1999).
2. Rico, F., Gonzalez, L., Casuso, I., Puig-Vidal, M. & Scheuring, S. High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science* **342**, 741–743 (2013).



#### 3.4 PUBLICATION P4: Biasing Effects of Receptor-Ligand Complexes on Protein-Unfolding Statistics

The following publication presents a model-free theoretical framework for the biasing effect on protein unfolding and receptor-ligand unbinding statistics in SMFS experiments. It predicts the change in force distribution caused by overlapping probability distributions of molecular fingerprints and pulling handles. We use Monte Carlo simulations to validate our findings by simulating the change in fingerprint unfolding behavior and receptor-ligand dissociation. A proposed method can be used for the extraction of corrected kinetic and energetic parameters from otherwise biased distributions.

# Biasing Effects of Receptor-Ligand Complexes on Protein-Unfolding Statistics

Constantin Schoeler, Tobias Verdorfer,  
Hermann E. Gaub and Michael A. Nash

*published in*

Physical Review E, 94(4), 042412, (2016)

*Reprinted from [116], with permission from American Physical Society  
Copyright 2015, American Physical Society*

**Biasing effects of receptor-ligand complexes on protein-unfolding statistics**Constantin Schoeler,<sup>1</sup> Tobias Verdorfer,<sup>1</sup> Hermann E. Gaub,<sup>1</sup> and Michael A. Nash<sup>1,2,3,\*</sup><sup>1</sup>*Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 Munich, Germany*<sup>2</sup>*Department of Chemistry, University of Basel, Klingelbergstrasse 80, 4056 Basel, Switzerland*<sup>3</sup>*Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich (ETH-Zürich), 4058 Basel, Switzerland*

(Received 30 May 2016; published 13 October 2016)

Protein receptor-ligand pairs are increasingly used as specific molecular handles in single-molecule protein-unfolding experiments. Further, known marker domains, also referred to as fingerprints, provide unique unfolding signatures to identify specific single-molecule interactions, when receptor-ligand pairs themselves are investigated. We show here that in cases where there is an overlap between the probability distribution associated with fingerprint domain unfolding and that associated with receptor-ligand dissociation, the experimentally measured force distributions are mutually biased. This biasing effect masks the true parameters of the underlying free energy landscape. To address this, we present a model-free theoretical framework that corrects for the biasing effect caused by such overlapping distributions.

DOI: [10.1103/PhysRevE.94.042412](https://doi.org/10.1103/PhysRevE.94.042412)**I. INTRODUCTION**

Mechanical forces play an important role in many biological systems. Refolding of individual titin domains is believed to assist in muscle contraction [1], stretching forces expose cryptic binding sites involved in focal adhesions [2], and mechanically stable receptor-ligand pairs govern the assembly of large extracellular machineries and adhesion of bacterial cells to their cellulosic carbon sources [3,4]. Single-molecule pulling experiments with atomic force microscopes [5], optical tweezers [6], or magnetic tweezers [7] have become widely used techniques to study such phenomena at the single-molecule level.

Due to the stochastic nature of domain unfolding, typical atomic force microscopy experiments record many thousands of data traces to obtain statistically meaningful results from single-molecule pulling experiments. To unambiguously identify the unfolding signals of a given protein domain of interest or the dissociation of a receptor-ligand system under external load, the resulting data sets need to be filtered, and specific, single-molecule interactions must be identified.

To accommodate this need, the community has developed an elegant strategy to achieve high yields of specific curves: In this approach, protein domains of interest are fused to a receptor complex that serves as a specific handle in pulling experiments. This improves curve yields and data fidelity by providing a specific molecular interaction handle to “grab” the protein of interest. Thereby, the unfolding of individual domains and the dissociation of a receptor-ligand complex can be studied in a single experiment [3,4,8–10]. Also, using a known protein domain in the fusion construct provides a unique unfolding pattern that can be used to identify specific traces, when receptor-ligand unbinding itself is studied. These domains are then referred to as fingerprint domains [5].

In order for a curve to be unambiguously identified as constituting specific signal, it needs to exhibit unfolding of all

included fingerprint domains plus a specific receptor-ligand dissociation event. Throughout this letter we will refer to domains fused to a receptor-ligand complex as fingerprint domains for both scenarios, namely protein-unfolding studies using receptor-ligand complexes as specific handles, as well as unbinding studies of receptor-ligand complexes of interest, which use fingerprint domains for assistance in data filtering. We discuss the statistical effects that arise when the respective force distributions for fingerprint domain unfolding and receptor-ligand complex rupture exhibit a finite overlap. We quantitatively show how the statistical overlap between these two distributions affects the experimentally observable unfolding and rupture force distributions. We provide a framework for extracting kinetic and energetic information from the experimentally observed distributions that are corrected for the biasing effects arising from such overlaps in a model-free fashion.

**II. THEORETICAL FRAMEWORK**

The standard theoretical framework treats protein unfolding or bond dissociation as thermally driven escape over a free energy barrier that is modulated by an external force  $F$  [11–14]. This description leads to a general expression for the distribution of unfolding or rupture forces in a pulling experiment,

$$p(F) = \frac{k(F)}{\dot{F}(F)} \exp\left(-\int_0^F dF' \frac{k(F')}{\dot{F}(F')}\right), \quad (1)$$

where  $k(F)$  is the force-dependent off rate of the system, and  $\dot{F}(F)$  is the force loading rate. In the simplest picture [12,13], the force-dependent off rate is given by [11]

$$k(F) = k_0 \exp\left(\frac{F\Delta x}{k_B T}\right) \quad (2)$$

where  $k_0$  and  $\Delta x$  are the zero-force off rate and distance to the free energy barrier, respectively. For a constant force loading rate  $\dot{F}$  and an off rate from Eq. (2), the integral in Eq. (1) can be solved analytically [Fig. 1(a)]. Dudko

\* michael.nash@unibas.ch

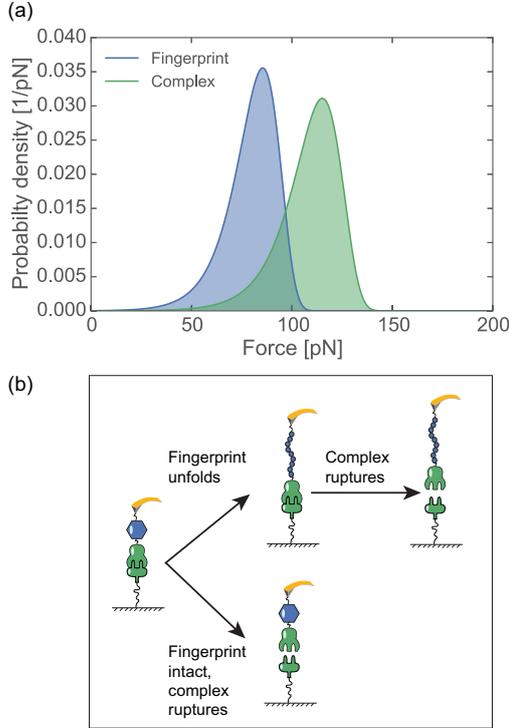


FIG. 1. (a) Overlapping distributions of fingerprint unfolding (blue region) and complex rupture (green region) for a constant loading rate  $\dot{F} = 200 \text{ pN s}^{-1}$  with fingerprint  $\Delta x = 0.4 \text{ nm}$ , and  $k_0 = 0.005 \text{ s}^{-1}$  and complex parameters  $\Delta x = 0.35 \text{ nm}$ , and  $k_0 = 0.001 \text{ s}^{-1}$  obtained by integrating Eq. (1) using Eq. (2). (b) Schematic of possible outcomes for the situation in (a). Analyzable data show fingerprint unfolding followed by complex rupture (upper path). Due to the overlapping distributions for unfolding and rupture, complex rupture with an intact fingerprint is also possible (lower path).

et al. [15] have used the Kramers theory [16] to obtain a more sophisticated expression for the force-dependent off rate, which accounts for the specific shape of the free energy landscape. This approach also provides an analytical solution to Eq. (1) for a constant loading rate and includes the height of the free energy barrier  $\Delta G$  as an additional parameter. Over the years, more theoretical models describing the distributions for domain unfolding and receptor-ligand dissociation have been developed [17–22] that can be applied to experimentally measured force distributions.

Since a constant force loading rate is required to obtain an analytical solution for the distribution of forces in a pulling experiment, force ramp measurements [23,24], where the external force is ramped linearly, are an elegant technique to study protein unfolding and receptor-ligand dissociation. In an experiment where a receptor-ligand system is used to probe the unfolding behavior of a protein fingerprint domain

of interest, care has to be taken when analyzing the resulting unfolding or rupture force distributions. If the probability distributions for fingerprint domain unfolding and complex rupture are disjunct, i.e., the complex ruptures at forces well above those required for fingerprint unfolding, the measured distributions are unbiased and can be readily analyzed using a desired form of Eq. (1). If those distributions have a substantial overlap [Fig. 1(a)], however, biasing effects have to be taken into account.

To pass the data analysis filter a given curve is required to exhibit both fingerprint unfolding and complex rupture [Fig. 1(b); upper path], i.e., the complex must not rupture prior to fingerprint unfolding [Fig. 1(b); lower path]. Consequently, the resulting distribution of fingerprint unfolding forces will be shifted downwards towards lower forces. By the same logic, experimentally observed distributions for receptor-ligand complex rupture forces will be truncated at the lower end and shifted upwards toward higher forces in a constant-loading-rate experiment. This biasing effect has been used qualitatively by Jobst et al. [8] to unambiguously identify a redundant dual binding mode in a receptor-ligand complex.

Here we treat this biasing effect quantitatively and calculate these effects independent of the model used in Eq. (1). For overlapping distributions of fingerprint,  $p_f(F)$ , and receptor-ligand complex,  $p_c(F)$ , the biased distribution of the fingerprint,  $p_f^*(F)$ , can be calculated by multiplying the original distribution by the cumulative probability for the bond to rupture at higher forces and renormalizing,

$$p_f^*(F) = \frac{p_f(F) \int_F^\infty dF' p_c(F')}{\eta}, \quad (3)$$

where  $\eta$  is a normalization constant. Since  $p_c(f)$  is normalized, Eq. (3) can be rewritten,

$$p_f^*(F) = \frac{p_f(F) \left(1 - \int_0^F dF' p_c(F')\right)}{\eta} \quad (4)$$

$$= \frac{p_f(F) (1 - P_c(F))}{\eta}, \quad (5)$$

where  $P_c(F)$  is the cumulative distribution function of the complex rupture forces. The normalization constant  $\eta$  can be calculated by integrating over the numerator in Eq. (5):

$$\eta = 1 - \int_0^\infty dF p_f(F) P_c(F). \quad (6)$$

Intuitively, the biased fingerprint distribution is normalized by the ratio of curves that exhibit fingerprint unfolding vs all rupture events. The above calculations apply for both constant-loading-rate and constant-speed measurements. By the same logic, the biased distribution of observed complex ruptures for a constant loading rate can be calculated as

$$p_c^*(F) = \frac{p_c(F) P_f(F)}{\eta}. \quad (7)$$

Both biased distributions for fingerprint unfolding,  $p_f^*(F)$ , and complex ruptures,  $p_c^*(F)$ , are normalized by the same yield parameter  $\eta$  since both distributions are extracted from the same curves in a given data set. For a mathematical proof, see Appendix A. We note that the biasing effect on complex rupture in the constant-speed case is more difficult to quantify.

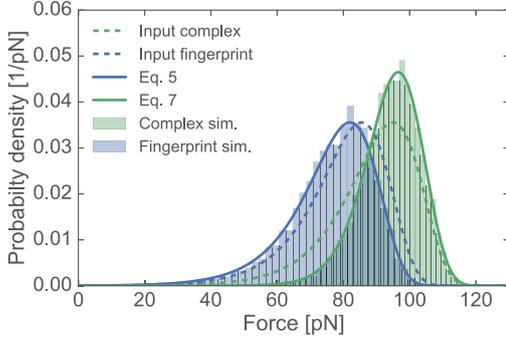


FIG. 2. Force ramp simulation with  $\dot{F} = 200 \text{ pN s}^{-1}$ , complex  $\Delta x_c = 0.4 \text{ nm}$ ,  $k_{0,c} = 0.002 \text{ s}^{-1}$ , and fingerprint  $\Delta x_f = 0.4 \text{ nm}$ ,  $k_{0,f} = 0.005 \text{ s}^{-1}$  ( $\eta = 0.71$ ). Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

Since the additional contour length released upon fingerprint unfolding is not immediately compensated for by a feedback on the force signal, there is a pronounced drop in force associated with fingerprint unfolding, giving rise to the characteristic sawtooth pattern in force extension traces. Ignoring the finite relaxation time of the harmonic pulling device, the force will drop from  $F_1 = F(x, L)$  to  $F_2 = F(x, L + \Delta L)$ , where the former describes the (nonlinear) elastic behavior of the pulling device and potential linker molecules, and  $\Delta L$  is the additional contour length released upon fingerprint unfolding. Whether or not the observed complex distribution is biased depends on the value of  $p_c(F_2)$ . For  $p_c(F_2) \approx 0$ , no biasing will occur, whereas  $p_c(F_2) > 0$  will cause a biasing effect. In practice,  $\Delta L$  is usually large enough to ensure that complex distributions are unbiased in constant-speed experiments, leaving a substantial bias only on the observed fingerprint distribution. A strategy to implement our correction procedure for the constant-speed protocol is proposed in Appendix B.

### III. MONTE CARLO SIMULATIONS

To validate our results, we used a Monte Carlo approach to simulate fingerprint domain unfolding in combination with receptor-ligand dissociation in a constant-loading-rate

TABLE I. Input vs fit parameters of the simulation shown in Fig. 2. For these parameters, the yield parameter equals  $\eta = 0.71$ . Initially, the simulated distributions were fit with the uncorrected distributions ( $k_{0,\text{fit}}$  and  $\Delta x_{\text{fit}}$ ). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions [Eqs. (5) and (7)] in a global fitting procedure to obtain the corrected parameters  $k_{0,\text{global}}$  and  $\Delta x_{\text{global}}$ .

	$k_{0,\text{input}} (\text{s}^{-1})$	$k_{0,\text{fit}} (\text{s}^{-1})$	$k_{0,\text{global}} (\text{s}^{-1})$	$\Delta x_{\text{input}} (\text{nm})$	$\Delta x_{\text{fit}} (\text{nm})$	$\Delta x_{\text{global}} (\text{nm})$
Fingerprint	$5.0 \times 10^{-3}$	$(6.6 \pm 0.2) \times 10^{-3}$	$(4.7 \pm 0.4) \times 10^{-3}$	0.400	$0.402 \pm 0.007$	$0.401 \pm 0.005$
Complex	$2.0 \times 10^{-3}$	$(0.10 \pm 0.02) \times 10^{-3}$	$(1.9 \pm 0.02) \times 10^{-3}$	0.400	$0.527 \pm 0.007$	$0.402 \pm 0.005$

protocol. Our simulation routine is similar to the approach described in Ref. [25] and uses the phenomenological model due to its analytical tractability. Briefly, we integrate Eq. (2) over a time step  $\Delta t$ , where  $F = F(t) = \dot{F}t$ , to obtain probabilities  $p_u$  and  $p_r$  for fingerprint unfolding and complex rupture, respectively. These probabilities are compared to independent random numbers  $n_u$  and  $n_r$  between 0 and unity. If  $p_u < n_u$  and  $p_r < n_r$ , the fingerprint and complex remain intact and the next iteration is started. If  $p_u > n_u$  and  $p_r < n_r$ , the corresponding force is noted as the fingerprint unfolding force and the next iteration is started with only the complex intact. The simulation then continues until  $p_r > n_r$  and the corresponding force is noted as the complex rupture force. If  $p_u < n_u$  and  $p_r > n_r$ , or  $p_u > n_u$  and  $p_r > n_r$ , the complex ruptured prior to fingerprint unfolding or at the same time, and an experimental curve would be unanalyzable and filtered out during the data analysis procedure.

Results from a Monte Carlo simulation at constant loading rate  $\dot{F} = 200 \text{ pN s}^{-1}$  with overlapping distributions for fingerprint unfolding and complex rupture are shown in Fig. 2.  $N = 10000$  traces were simulated and the observed fingerprint domain unfolding forces and receptor-ligand complex rupture forces were histogrammed. We only analyzed curves that showed both fingerprint unfolding and complex rupture to mimic experimental conditions. As expected, both fingerprint unfolding and complex rupture distributions (blue and green histograms in Fig. 2) are shifted from the unbiased input distributions (dashed blue and green lines in Fig. 2). The biased results are well described by our theoretical predictions, which are shown as solid blue and green lines in Fig. 2. To illustrate the potential errors that can occur from not accounting for the fingerprint biasing effect, we used the uncorrected distribution [Eq. (1)] to fit the biased fingerprint domain unfolding and complex rupture histograms and compared the resulting fit parameters  $\Delta x_{\text{fit}}$  and  $k_{0,\text{fit}}$  with the unbiased input parameters (Table I). We found that for fingerprint domain unfolding,  $\Delta x$  is hardly affected, while  $k_0$  is overestimated by 30%. For complex rupture  $\Delta x$  is overestimated by 29% due to the smaller width of the biased distribution, while  $k_0$  is underestimated by over an order of magnitude. If the unbiased parameters for the complex distribution are known from a control experiment, our predicted biased distributions can be used to fit the experimental data to obtain unbiased values for the fit parameters pertaining to the fingerprint, or vice versa. Alternatively, a global fitting procedure can be applied to both biased distributions for constant-loading-rate data to obtain unbiased fit parameters without prior knowledge of either distribution. Using this approach, we obtained global fit parameters that were within 6% of the input parameters (Table I).

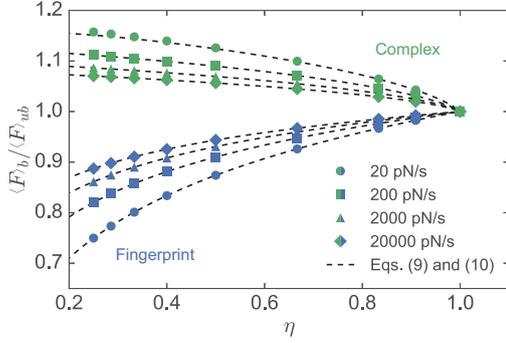


FIG. 3. Biasing of fingerprint unfolding and complex rupture forces as a function of the theoretical yield of curves exhibiting fingerprint unfolding  $\eta$ . Data points show the relative mean biased unfolding force  $\langle F \rangle_b / \langle F \rangle_{ub}$ ; fingerprint and complex data are shown in blue and green, respectively. For fingerprint data, parameters were held constant at  $\Delta x_f = 0.4$  nm and  $k_{0,f} = 0.005$  s $^{-1}$ , while the complex distribution was shifted by maintaining  $\Delta x_c = 0.4$  nm and varying the off rate. For complex data,  $\Delta x_c = 0.4$  nm and  $k_{0,c} = 0.005$  s $^{-1}$  were held constant and the fingerprint distribution was shifted by maintaining  $\Delta x_f = 0.4$  nm and varying the off rate. Dashed lines represent predictions based on Eq. (9) and (10).

#### IV. SPECIAL CASE: EQUAL POTENTIAL WIDTHS

In Fig. 3 we quantify the magnitude of the biasing effect by numerically calculating the relative mean biased unfolding force of the fingerprint (blue symbols) and complex (green symbols)  $\langle F \rangle_b / \langle F \rangle_{ub} = \int dF F p_{\tilde{v}_c}^b(F) / \int dF F p_{\tilde{v}_c}(F)$  as a function of the theoretical ratio of curves exhibiting fingerprint unfolding prior to receptor-ligand complex rupture,  $\eta$  [Eq. (6)]. For analytical tractability we chose the special case  $\Delta x_c = \Delta x_f$ . With this simplification we find for the fingerprint

$$\langle F \rangle_b = \frac{k_B T}{\Delta x_f} e^{\frac{k_B T k_{0,f}}{\Delta x_f r}} E_1 \left( \frac{k_B T k_{0,f}}{\Delta x_f r \eta} \right), \quad (8)$$

where  $E_1(x)$  is the exponential integral. Using  $e^x E_1(x) \cong \ln(1 + \frac{e^{-x}}{x})$  we can produce an analytical approximation for the relative mean biased unfolding force for the aforementioned special case,

$$\frac{\langle F \rangle_b}{\langle F \rangle_{ub}} \cong \frac{\ln(1 + \frac{r \Delta x_f}{k_{0,f} k_B T} e^{-\gamma} \eta)}{\ln(1 + \frac{r \Delta x_f}{k_{0,f} k_B T} e^{-\gamma})}, \quad (9)$$

where  $\gamma = 0.577 \dots$  is the Euler-Mascheroni constant. The analogous result for the complex reads

$$\frac{\langle F \rangle_b}{\langle F \rangle_{ub}} \cong \frac{1}{\eta} - \frac{\ln(1 + \frac{r \Delta x_c}{k_{0,c} k_B T} e^{-\gamma} (1 - \eta))}{\ln(1 + \frac{r \Delta x_c}{k_{0,c} k_B T} e^{-\gamma})} \frac{1 - \eta}{\eta}. \quad (10)$$

For this special case Eq. (6) can be evaluated analytically, yielding  $\eta = (1 + \frac{k_{0,c}}{k_{0,f}})^{-1}$ . Equations (9) and (10) (dashed lines in Fig. 3) agree very well with our numerical results over the loading-rate regime covered. Figure 3 clearly shows that the biasing effect is more pronounced for low loading rates, consistent with our theoretical predictions based on

Eqs. (9) and (10). In cases where data cannot be corrected for a potential biasing effect, e.g., due to low experimental yields, the magnitude of the biasing effect can be minimized by increasing the loading rate in a pulling experiment.

#### V. CONCLUSION

Our calculations provide a framework for analyzing single-molecule force spectroscopy data where receptor-ligand systems are used as specific handles to study a fingerprint domain of interest, or vice versa. In such experiments, it is many times the case that the distributions of fingerprint domain unfolding and complex rupture have a significant overlap (a few exemplary cases can be found in Refs. [3,9,10,26]). In this case biasing effects will occur and should be considered in the analysis procedure. Our findings can be applied to both constant-speed and force ramp (constant-loading-rate) experimental protocols, however, it should be noted that the biasing effect on complex unbinding is more complicated in the constant-speed protocol, due to the drop in force upon fingerprint unfolding dependent on the length of the unfolded domain. Since the biasing effects are solely due to the statistical nature of domain unfolding or complex unbinding, our results, specifically Eqs. (5) and (7), are valid regardless of the specific model used in Eq. (1).

#### ACKNOWLEDGMENTS

The authors would like to thank Ellis Durner, Markus A. Jobst, Magnus S. Bauer, Fabian Baumann, and Steffen Sedlak for valuable discussions. This work was supported by an advanced grant to H.E.G. from the European Research Council (Cellufuel Grant No. 294438). M.A.N. acknowledges funding from Society in Science The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland.

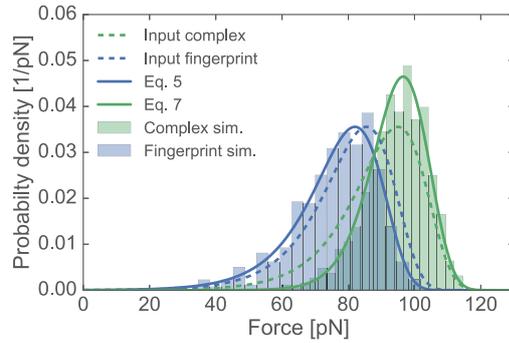


FIG. 4. Force ramp simulation with  $\dot{F} = 200$  pN s $^{-1}$ , complex  $\Delta x_c = 0.4$  nm,  $k_{0,c} = 0.002$  s $^{-1}$ , fingerprint  $\Delta x_f = 0.4$  nm,  $k_{0,f} = 0.005$  s $^{-1}$  ( $\eta = 0.71$ ), and simulated traces  $N = 1000$ . Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

TABLE II. Input vs fit parameters of the simulation shown in Fig. 4. For these parameters, the yield parameter equals  $\eta = 0.71$ . Initially, the simulated distributions were fit with the uncorrected distributions ( $k_{0,\text{fit}}$  and  $\Delta x_{\text{fit}}$ ). To correct for the biasing effects, both fingerprint and complex data were fit with their respective biased distributions [Eqs. (5) and (7)] in a global fitting procedure to obtain the corrected parameters  $k_{0,\text{global}}$  and  $\Delta x_{\text{global}}$ .

	$k_{0,\text{input}} \text{ (s}^{-1}\text{)}$	$k_{0,\text{fit}} \text{ (s}^{-1}\text{)}$	$k_{0,\text{global}} \text{ (s}^{-1}\text{)}$	$\Delta x_{\text{input}} \text{ (nm)}$	$\Delta x_{\text{fit}} \text{ (nm)}$	$\Delta x_{\text{global}} \text{ (nm)}$
Fingerprint	$5.0 \times 10^{-3}$	$(5.9 \pm 1.2) \times 10^{-3}$	$(4.3 \pm 1.5) \times 10^{-3}$	0.400	$0.409 \pm 0.011$	$0.410 \pm 0.020$
Complex	$2.0 \times 10^{-3}$	$(0.2 \pm 0.1) \times 10^{-3}$	$(2.7 \pm 1.3) \times 10^{-3}$	0.400	$0.504 \pm 0.020$	$0.384 \pm 0.021$

#### APPENDIX A: IDENTITY OF YIELD PARAMETER $\eta$ IN EQS. (5) AND (7)

As we state in the text, the normalization parameter  $\eta$  is equal to the ratio of curves that exhibit fingerprint unfolding vs all rupture events. In other words a fraction  $1 - \eta$  of all curves will be “missed” in an actual experiment, since they do not exhibit fingerprint unfolding and would hence be discarded during data analysis. Consequently, both fingerprint and complex distributions (which are obtained from the same curves) need to be normalized by the same  $\eta$  to obtain probability distributions normalized to unity. Mathematically,

one needs to show that

$$1 - \int_0^\infty dF p_f(F) P_c(F) = \eta = \int_0^\infty dF p_c(F) P_f(F). \quad (\text{A1})$$

We use integration by parts and  $\int_0^F dF' p(F') = P(F)$  to evaluate the right-hand side (rhs) of Eq. (A1):

$$\text{rhs} = P_c(F) P_f(F) \Big|_0^\infty - \int_0^\infty dF P_c(F) p_f(F). \quad (\text{A2})$$

Since  $P(0) = 0$  and  $P(\infty) = 1$  this is equal to the left-hand side of Eq. (A1).

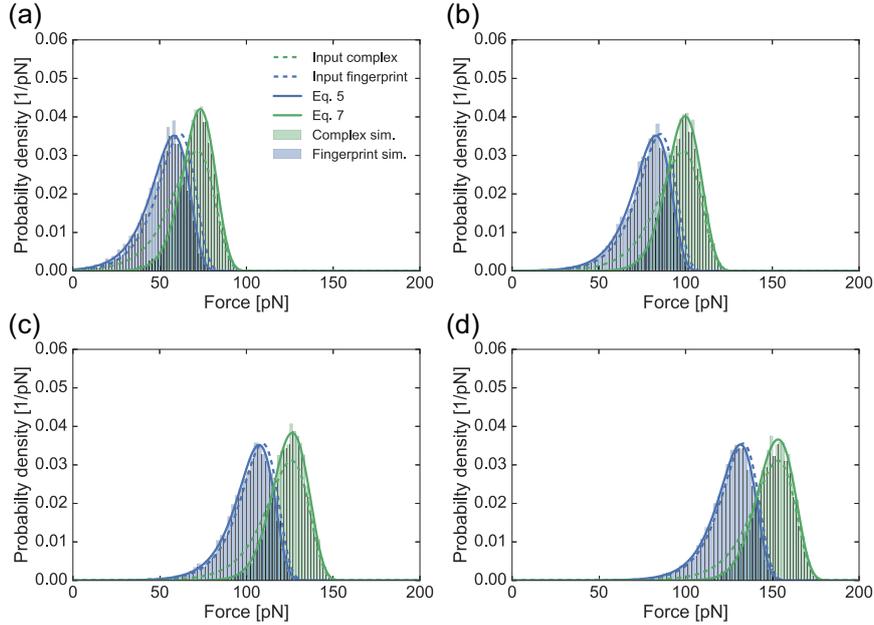


FIG. 5. Force ramp simulation with complex  $\Delta x_c = 0.35 \text{ nm}$ ,  $k_{0,c} = 0.004 \text{ s}^{-1}$ , fingerprint  $\Delta x_f = 0.4 \text{ nm}$ ,  $k_{0,f} = 0.005 \text{ s}^{-1}$ , and varying  $\dot{F}$ . Due to the different potential widths  $\Delta x$ , the yield parameter  $\eta$  changes with the loading rate. Values of the loading rate and resulting yield parameter are (a)  $\dot{F} = 20 \text{ pN s}^{-1}$  and  $\eta = 0.70$ , (b)  $\dot{F} = 200 \text{ pN s}^{-1}$  and  $\eta = 0.75$ , (c)  $\dot{F} = 2000 \text{ pN s}^{-1}$  and  $\eta = 0.80$ , and (d)  $\dot{F} = 20000 \text{ pN s}^{-1}$  and  $\eta = 0.85$ . Histograms of the simulated fingerprint unfolding forces and complex ruptures are shown in blue and green, respectively. Dashed blue and green lines represent the unbiased fingerprint unfolding and complex rupture force distributions, respectively. Biased unfolding and complex rupture force distributions for the fingerprint and complex calculated from Eqs. (5) and (7) are shown by solid blue and green lines, respectively.

## APPENDIX B: CONSTANT SPEED

As pointed out in the text, the correction procedure for distributions obtained from constant-speed measurements is more involved. In this Appendix we discuss a strategy for extracting unbiased parameters from fingerprint distributions for this pulling protocol. Due to the nonlinear elasticity of linker molecules (e.g., polyethylene glycol, spacers, or unfolded protein backbone), the force loading rate  $\dot{F} = \dot{F}(F)$  becomes a function of the force and the integral in Eq. (1) can no longer be evaluated analytically. In a standard pulling experiment, a harmonic pulling device (e.g., atomic force microscopy cantilever or optically trapped bead) is connected to the aforementioned linker molecules. By applying a force balance it can be shown that the force loading rate is given by [14]

$$\frac{v}{\dot{F}(F)} = \frac{1}{k_h} + \frac{\partial x(F)}{\partial F}, \quad (\text{B1})$$

where  $v$  is the pulling speed,  $k_h$  is the spring constant of the harmonic pulling device, and  $x(F)$  is the force-dependent extension of the linker. The biased distribution for fingerprint unfolding  $p_i^*(F)$  can be computed by numerically solving the integrals occurring in Eqs. (5) and (6), using a model for the elastic response of the linker molecules such as the worm-like-chain model, freely rotating model, or composite

model proposed by Livadaru et al. [27] in Eq. (B1). The choice of model should be made based on the force regime in which fingerprint unfolding and complex rupture are expected and the molecular linkers present in an experimental setup need to be accounted for in these models via their contour length,  $L$ , and elasticity, e.g., persistence length  $p$ . The force-dependent loading rate for the worm-like-chain model has been derived as Eq. (4) in Ref. [14].

## APPENDIX C: SUPPLEMENTAL FIGURES AND TABLES

This Appendix contains two figures and one table that support the results in the text. Figure 4 shows the results of a simulation with parameters identical to those in Fig. 2, except only  $N = 1000$  traces were simulated, to mimic a total number of curves more similar to average experimental yields. The extracted fit parameters for this simulation are listed in Table II. We note that despite the increase in uncertainties, the extracted parameters from our global fitting procedure still reproduce the input parameters much better than those obtained from a fit to the uncorrected distributions. Figure 5 shows the results of Monte Carlo simulations at different loading rates ranging from  $\dot{F} = 20 \text{ pN s}^{-1}$  to  $\dot{F} = 20000 \text{ pN s}^{-1}$ . Due to the different potential widths  $\Delta x_c = 0.35 \text{ nm}$  and  $\Delta x_f = 0.4 \text{ nm}$ , the yield parameter  $\eta$  varies for the different simulations.

- 
- [1] J. A. Rivas-Pardo, E. C. Eckels, I. Popa, P. Kosuri, W. A. Linke, and J. M. Fernández, *Cell Reports* **14**, 1339 (2016).
- [2] A. del Rio, R. Perez-Jimenez, R. Liu, P. Roca-Cusachs, J. M. Fernandez, and M. P. Sheetz, *Science* **323**, 638 (2009).
- [3] S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, and H. E. Gaub, *Proc. Natl. Acad. Sci. USA* **109**, 20431 (2012).
- [4] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, *Nature Commun.* **5**, 5635 (2014).
- [5] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, *J. Struct. Biol.* (2016), doi: 10.1016/j.jsb.2016.02.011.
- [6] K. C. Neuman and A. Nagy, *Nature Methods* **5**, 491 (2008).
- [7] S. B. Smith, L. Finzi, and C. Bustamante, *Science* **258**, 1122 (1992).
- [8] M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, *eLife* **4**, e10319 (2015).
- [9] F. Baumann, M. S. Bauer, L. F. Milles, A. Alexandrovich, H. E. Gaub, and D. A. Pippig, *Nature Nanotechnol.* **11**, 89 (2016).
- [10] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub, *Nature Methods* **11**, 1127 (2014).
- [11] G. Bell, *Science* **200**, 618 (1978).
- [12] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, *Biophys. J.* **72**, 1568 (1997).
- [13] E. Evans and K. Ritchie, *Biophys. J.* **72**, 1541 (1997).
- [14] O. K. Dudko, G. Hummer, and A. Szabo, *Proc. Natl. Acad. Sci. USA* **105**, 15755 (2008).
- [15] O. K. Dudko, G. Hummer, and A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
- [16] H. A. Kramers, *Physica* **7**, 284 (1940).
- [17] O. K. Dudko, *Q. Rev. Biophys.* **49**, e3 (2015).
- [18] B. Heymann and H. Grubmüller, *Phys. Rev. Lett.* **84**, 6126 (2000).
- [19] J. T. Bullerjahn, S. Sturm, and K. Kroy, *Nature Commun.* **5**, 4463 (2014).
- [20] R. W. Friddle, *Phys. Rev. Lett.* **100**, 138302 (2008).
- [21] R. W. Friddle, A. Noy, and J. J. De Yoreo, *Proc. Natl. Acad. Sci. USA* **109**, 13573 (2012).
- [22] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, *Phys. Rev. Lett.* **94**, 048301 (2005).
- [23] A. F. Oberhauser, P. K. Hansma, M. Carrión-Vázquez, and J. M. Fernandez, *Proc. Natl. Acad. Sci. USA* **98**, 468 (2001).
- [24] E. Evans, A. Leung, V. Heinrich, and C. Zhu, *Proc. Natl. Acad. Sci. USA* **101**, 11281 (2004).
- [25] C. Friedsam, A. K. Wehle, F. Kühner, and H. E. Gaub, *J. Phys.: Condens. Matter* **15**, S1709 (2003).
- [26] J. P. Mueller, S. Mielke, A. Loeff, T. Obser, C. Beer, L. K. Bruetzel, D. A. Pippig, W. Vanderlinden, J. Lipfert, R. Schneppenheim, and M. Benoit, *Proc. Natl. Acad. Sci. USA* **113**, 1208 (2016).
- [27] L. Livadaru, R. R. Netz, and H. J. Kreuzer, *Macromolecules* **36**, 3732 (2003).





### 3.5 PUBLICATION P5: Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity

The following publication presents a highly sensitive assay for the detection and spatial localization of cellulolytic enzyme activity. The described hydrogel reagent signaling system is based on a redox-initiated polymerization reaction, which localizes the crosslinking of a fluorescent hydrogel to locations of cellulose hydrolysis. This system allows the quantification of the activity of enzyme cocktails on various substrates in a rapid parallel format, and real-time imaging of the biomass degradation process using time-lapse TIRF microscopy and AFM imaging.

My contribution to this work was the development and calibration of the bulk assay, and data acquisition, analysis and interpretation for the bulk enzyme activity detection experiments.

# Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity

Klara H. Malinowska\*, Tobias Verdorfer\*, Aylin Meinhold,  
Lukas F. Milles, Victor Funk, Hermann E. Gaub and Michael A. Nash

*\*these authors contributed equally to this publication*

*published in*

ChemSusChem, 7(10), 2825 – 2831, (2014)

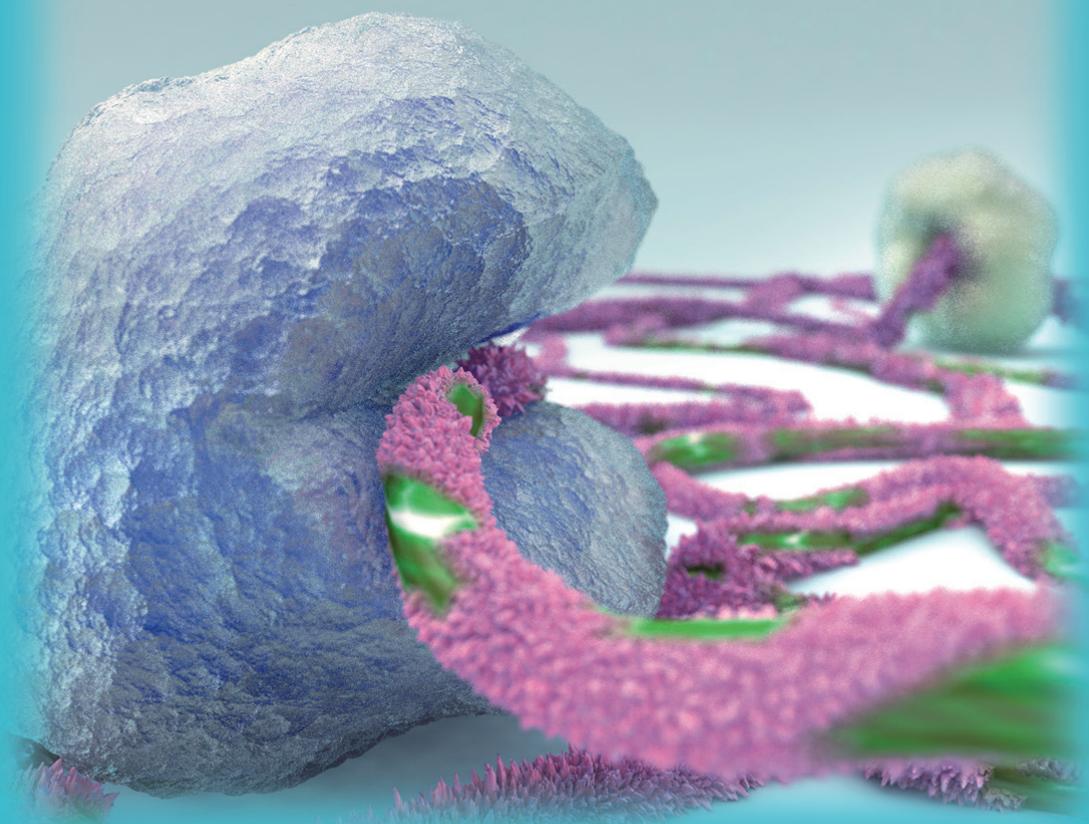
*Reprinted from [117], with permission from John Wiley and Sons  
Copyright 2014, John Wiley and Sons*

ISSN 1864-5631 · Vol. 7 · No. 10 · October, 2014

CHEMISTRY & SUSTAINABILITY

# CHEM**SUS**CHEM

ENERGY & MATERIALS



A Journal of



ChemPubSoc  
Europe

Supported by  
**ACES**

10/2014

[www.chemsuschem.org](http://www.chemsuschem.org)

**Cover:** Hydrogel System for Detection and Imaging of Enzyme Activity (M. A. Nash)

**Review:** Sonochemistry for Conversion of Biomass to Platform Chemicals (G. Chatel)

**Highlight:** Oxidant-Free Dehydrogenative Coupling via Hydrogen Evolution (Y. Li)

**Communication:** Hydrogenated Furanes as Biomass-Based Diesel Additives (A. Bell)

WILEY-VCH

DOI: 10.1002/cssc.201402428

## Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity

Klara H. Malinowska, Tobias Verdorfer, Aylin Meinhold, Lukas F. Milles, Victor Funk, Hermann E. Gaub, and Michael A. Nash<sup>\*[a]</sup>

Understanding the process of biomass degradation by cellulolytic enzymes is of urgent importance for biofuel and chemical production. Optimizing pretreatment conditions and improving enzyme formulations both require assays to quantify saccharification products on solid substrates. Typically, such assays are performed using freely diffusing fluorophores or dyes that measure reducing polysaccharide chain ends. These methods have thus far not allowed spatial localization of hydrolysis activity to specific substrate locations with identifiable morphological features. Here we describe a hydrogel reagent signaling (HyReS) system that amplifies saccharification products and initiates crosslinking of a hydrogel that localizes to locations of

cellulose hydrolysis, allowing for imaging of the degradation process in real time. Optical detection of the gel in a rapid parallel format on synthetic and natural pretreated solid substrates was used to quantify activity of *T. emersonii* and *T. reesei* enzyme cocktails. When combined with total internal reflection fluorescence microscopy and AFM imaging, the reagent system provided a means to visualize enzyme activity in real-time with high spatial resolution (<2  $\mu\text{m}$ ). These results demonstrate the versatility of the HyReS system in detecting cellulolytic enzyme activity and suggest new opportunities in real-time chemical imaging of biomass depolymerization.

### Introduction

Multistep bioconversion processes for production of liquid fuels and other chemical commodities from biomass are poised to alter our energy future. One step on the route to biomass-derived fuels is the enzymatic hydrolysis of cellulosic materials into fermentable sugars, a keystone in the overall process.



COVER

Cellulolytic enzymes are used in large quantities to depolymerize cellulose chains into energy-dense glucose monomers and other short chain cellodextrins prior to fermentation.<sup>[1]</sup> In order to achieve high conversion rates in practice, enzymatic saccharification requires high enzyme loadings (e.g., 20 mg enzyme  $\text{g}_{\text{substrate}}^{-1}$ ) and can be costly and inefficient.<sup>[2]</sup> To make the process more efficient and affordable, pretreatment methods that render the substrate more susceptible to enzymatic degrada-

tion have been developed.<sup>[3]</sup> Additionally, enzyme cocktails secreted from the aerobic fungus *Trichoderma reesei* (*Tr*) are being steadily improved to exhibit synergism among components for industrial processes.<sup>[4]</sup> This continued improvement has meanwhile drawn attention to a major challenge in the field, namely that of assaying and quantifying the effectiveness of cellulolytic enzyme formulations on a range of substrates possessing variable composition, morphology, degrees of crystallinity, and/or lignin content.

In the past, cellulase assays have been performed using a suite of bulk biochemical methods.<sup>[5–13]</sup> These include a variety of assays which measure the content of reducing polysaccharide chain ends using redox-sensitive absorbing dyes [e.g., 3,5-dinitrosalicylic acid (DNS)]. Other methods include the glucose oxidase (GOx)/horseradish peroxidase system (HRP)<sup>[14]</sup> which provides a fluorescent readout, or HPLC combined with quantitative mass spectrometry,<sup>[15]</sup> which reports on the quantity and size distribution of hydrolyzed chains. Electrochemical biosensors have also been employed to detect cellulase activity.<sup>[16]</sup>

More recently, methods for observing the spatial localization of cellulolytic activity have garnered interest as well. Imaging substrate locations susceptible to enzymatic hydrolysis could allow correlation of digestibility with substrate features such as fiber bundle size, degree of fiber branching, and/or crystal orientation. Conventional high-resolution imaging methods (e.g., TEM, SEM) are performed under vacuum and therefore are not suitable for monitoring enzymatic digestibility under biocompatible conditions. AFM imaging in liquid has shown promise, and has been used to observe disintegration of ultraflat micro-

[a] K. H. Malinowska,<sup>\*</sup> T. Verdorfer,<sup>\*</sup> A. Meinhold, L. F. Milles, V. Funk, Prof. Dr. H. E. Gaub, Dr. M. A. Nash  
Lehrstuhl für Angewandte Physik and Center for Nanoscience  
Ludwig-Maximilians-Universität  
Amalienstrasse 54, 80799 Munich (Germany)  
E-mail: michael.nash@lmu.de

<sup>†</sup> These authors contributed equally to this work.

Supporting Information for this article is available on the WWW under <http://dx.doi.org/10.1002/cssc.201402428>.

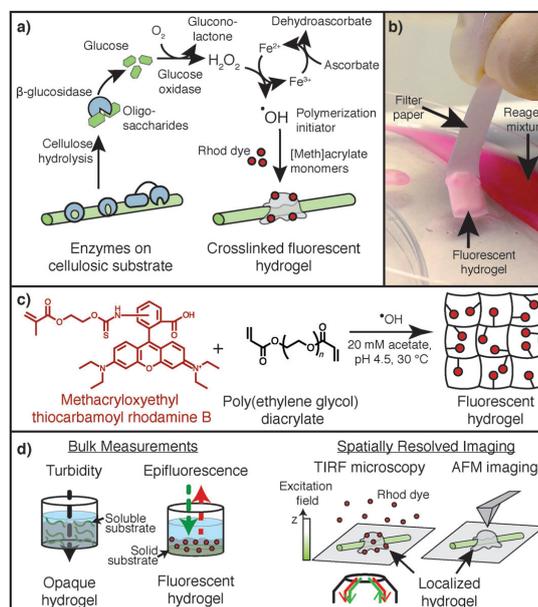
tomed substrates.<sup>[17,18]</sup> Time-resolution using AFM imaging is limited by scan times of up to several minutes and substrates are limited to ultraflat artificial cellulose surfaces (i.e., no native fibrils). Stimulated Raman spectroscopy has also been shown to provide adequate spatial and temporal resolution<sup>[19]</sup> and can be used on natural biomass substrates, however it is technically involved, requiring synchronization of multiple lasers at different wavelengths with modulation in the MHz range. Single-molecule fluorescence has shown potential for providing insights into cellulolytic enzyme function, but, so far, studies have only focused on carbohydrate binding modules and their cooperativity,<sup>[20–22]</sup> and the method has not been used to directly detect cellulolytic enzyme activity. Typically, soluble fluorescent enzyme substrates will diffuse away too quickly to allow for localization of activity. A fluorescent reagent system that could be used to directly read hydrolysis activity in an imaging modality could provide new insights to enzymatic activity and synergy.

Since its discovery in the late 19th century, hydroxyl radicals produced via Fenton chemistry have found use in many industrial applications, ranging from removal of organics from contaminated wastewater,<sup>[23]</sup> to redox-initiated free radical polymerization.<sup>[24,25]</sup> More recently in the biomaterials field, Fe<sup>II</sup> Fenton reagents have been combined with GOx to achieve spatially controlled release of hydroxyl radicals from pre-existent poly(ethylene glycol) (PEG) hydrogels. For example, spatial confinement of radical generation at an interface was used to prepare multilayer particles.<sup>[26]</sup> Fluorescent gels could also be produced in response to immuno-recognition events.<sup>[27–29]</sup>

Here we extend the use of Fe<sup>II</sup> Fenton reagents, and demonstrate their application in a cellulase-mediated polymerization system capable of monitoring cellulose hydrolysis in real time. The hydrogel reagent signaling system (HyReS system) described here detects cellulolytic enzyme activity with good sensitivity and is compatible with a variety of readout formats, including bulk turbidity and fluorescence as well as spatially-resolved total internal reflection fluorescence (TIRF) and AFM imaging, as depicted in Figure 1 d. The HyReS system relies on an Fe<sup>II</sup> Fenton reagent that is oxidized by hydrogen peroxide with concomitant production of a reactive hydroxyl radical.

## Results and Discussion

An overview of the HyReS system is shown in Figure 1 a. We used enzyme formulations that incorporated the synergistic endo- and exoglucanase activities of cellulolytic enzymes together with the cellobiase activity of  $\beta$ -glucosidase.  $\beta$ -glucosidase is frequently supplemented into cellulolytic enzyme formulations to convert cellobiose to glucose, thereby removing a primary inhibitor of exoglucanases in the cocktail.<sup>[30]</sup> In our system,  $\beta$ -glucosidase is responsible for production of glucose,

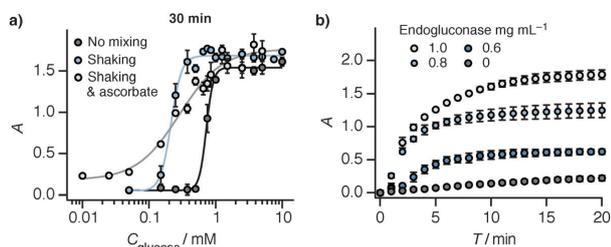


**Figure 1.** Overview of hydrogel reagent signaling (HyReS) system for detecting and imaging the degradation of cellulosic substrates. a) Saccharification products are converted into  $H_2O_2$  via reaction with  $\beta$ -glucosidase and GOx.  $H_2O_2$  proceeds with an  $Fe^{2+}$ -Fenton reagent to produce hydroxyl radicals that initiate hydrogel crosslinking. b) Photograph of filter paper partially submerged in the HyReS mixture for 30 min. c) Scheme showing structures of Rhod dye and gel cross-linker PEG diacrylate. d) Detection of the hydrogel using bulk measurements and spatially resolved imaging. Left: Bulk measurements in a parallel 96-well format provide a method for screening substrate pretreatment conditions or optimizing enzyme formulations on soluble and solid substrates. Right: High-resolution imaging methods such as TIRF microscopy and AFM-imaging allow detection of gel formation locally on fiber surfaces.

which is further oxidized by GOx, directly producing  $H_2O_2$ , a reactant in the Fenton reaction. Gel formation proceeded via hydroxyl radical initiated polymerization of PEG diacrylate in the mixture, as depicted in Figure 1 c. Figure 1 b shows a representative gel film that polymerized onto a piece of filter paper upon partial submersion into the HyReS system containing  $1 \text{ mg mL}^{-1}$  Tr enzyme cocktail for 30 min. The composition of the HyReS mixture can be found in Table 1.

**Table 1.** Composition of the HyReS system.

Component	Concentration
glucose oxidase	$1 \text{ mg mL}^{-1}$
$FeSO_4$	$250 \mu\text{M}$
ascorbic acid	$250 \mu\text{M}$
PEG diacrylate ( $M_n$ , 575)	15 wt%
acetate buffer, pH 4.5	20 mM
rhodamine B methacrylate	$3.5 \mu\text{M}$ (epifluorescence)/35 nM (TIRF)/none (turbidity, AFM)
cellulolytic enzymes	0–2 $\text{mg mL}^{-1}$



**Figure 2.** Detection of hydrogel polymerization by turbidity measurements on soluble substrates. a) Glucose standards were added to the HyReS system in a 96-well plate format. Absorbance at 600 nm due to light scattering by the polymerized hydrogel was measured after 30 min. Fits were performed using the Hill equation. b) Varying amounts of endoglucanase were added to CMC and the HyReS system. Turbidity was monitored over time. Gel polymerization proceeded proportional to CMCase activity of the enzyme and could be followed continuously in real time.

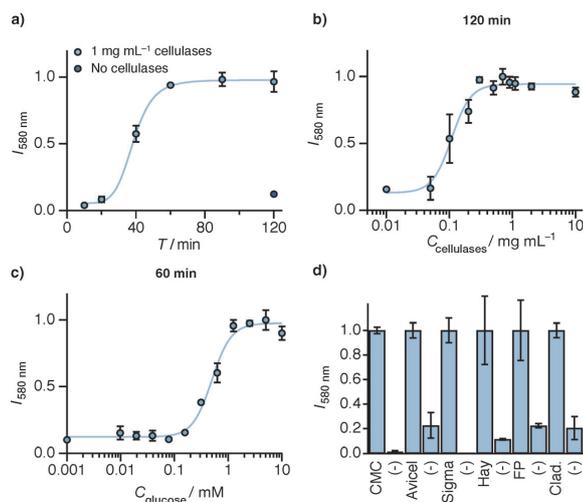
Initially, we tested the sensitivity of the HyReS system in detecting glucose directly added to sample wells of a 96-well plate. Since the PEG hydrogel turned the solution turbid as it polymerized, the absorbance signal at 550 nm increased with the amount of glucose in the solution. The results from a glucose standard curve measured after 30 min reaction time are shown in Figure 2a. A glucose detection limit in the low micromolar range was found. This sensitivity is similar to that found for microtiter plate DNS assays<sup>[31]</sup> and is generally sufficient for assaying cellulases involved in biomass conversion. Improvement in sensitivity was achieved by rotary shaking of the plate during the reaction. Inclusion of ascorbic acid in a 1:1 molar ratio with  $\text{FeSO}_4$  also improved the sensitivity by serving as a weak reducing agent in the HyReS system, reducing  $\text{Fe}^{\text{III}}$  back to  $\text{Fe}^{\text{II}}$ , thereby regenerating the Fenton catalyst in situ.<sup>[32]</sup> When using the standard HyReS system (Table 1) for detecting glucose, the dynamic range of detection was from 0.05 to 5 mM (Figure 2a).

Figure 2b shows an endoglucanase assay performed on the soluble cellulose analog carboxymethyl cellulose (CMC). Varying amounts of  $\beta$ -1,4-endoglucanase from the thermophilic fungus *Talaromyces emersonii* were added to 30 mM solutions of CMC and the HyReS system at 37 °C (without ascorbic acid in this case). Turbidity increased with CMCase activity in a concentration dependent manner. Interestingly, the final absorbance values achieved by different endoglucanase concentrations were not the same, suggesting the kinetics of polymerization affect the final absorbance signal generated. This result was likely attributable to differences in gel density which led to different optical extinction properties, or alternatively due to entrapment of the endoglucanase during hydrogel polymerization that restricted access to the CMC substrate.

Although CMC is commonly used for screening endoglucanase activity, it is a poor predictor of hydrolysis performance on pretreated natural biomass in the

context of biofuel production. For this purpose, solid substrates are typically more informative. To demonstrate the capabilities of the HyReS system on relevant solid substrates, hydrolysis on a variety of solid substrates was measured using fluorescence detection. Initially, Whatman #1 filter paper (FP) was used as the source of glucose. FP was cut into 6 mm disks and placed into the wells of a 96-well plate. The HyReS system including a fluorescent rhodamine monomer (Rhod) was added to the FP disks, along with 1 mg mL<sup>-1</sup> of *Tr*

enzymes. At given time points, the wells were washed to remove unreacted dye molecules, and the fluorescence was measured (Figure 3A). The result after 120 min was a pink-colored gel that conformally coated the filter paper, observable by eye with macroscopic dimensions (several mm thick). When the reagent system was added in the absence of the hydrolytic enzymes, background fluorescence remained low, indicating



**Figure 3.** Detection of polymerization by Rhod fluorescence on solid substrates. a) Rhod fluorescence intensity vs. time for HyReS system/*Tr* enzyme cocktail on filter paper. Samples were rinsed and fluorescence signal read at given time points (dark blue circle, lacking *Tr* enzymes). Hill equation fits serve as a guide for the eye. b) Fluorescence intensity vs. *Tr* enzyme concentration measured on filter paper after 120 min. c) Glucose standard for solid substrate. Small volumes of glucose standards were applied onto the filter paper to ensure similar diffusion geometry as during enzymatic hydrolysis of the substrate. HyReS system without cellulases was applied and fluorescence intensity was measured after 60 min. d) HyReS system/*Tr* enzymes were applied to cellululosic substrates for 2 h. Normalized signal was robust in comparison with negative controls. CMC: carboxymethyl cellulose; Avicel:  $\mu$ -crystalline cellulose; Sigma:  $\mu$ -crystalline cellulose powder; Hay: dilute acid pretreated hay; FP: filter paper; Clad.: pretreated algal *Cladophora* cellulose.

that the hydrogel assay was specific. Figure 3b shows the fluorescence signal after 120 min exposure of the HyReS system with varying concentration of *Tr* cellulases to the filter paper. These data show that our detection method discriminates between different levels of cellulolytic activity, with a linear dynamic range for *Tr* enzyme cocktails from 0.05 to 0.3 mg mL<sup>-1</sup>. The assay precision ranged from 2.0% at 0.3 mg mL<sup>-1</sup> *Tr* enzymes to 52% at 0.05 mg mL<sup>-1</sup> *Tr* enzymes.

To assay the absolute amount of glucose produced by cellulolytic enzymes on FP and not only the relative changes in activity, we calibrated the assay to glucose standards. To mimic the geometry of sugar release, FP disks were soaked with small volumes of concentrated glucose solutions in varying concentrations. The HyReS system including Rhod but lacking *Tr* enzymes was then added and samples were incubated for 1 h. Following rinsing, the fluorescence was measured (Figure 3c). The dynamic range of this standardization assay on glucose was found to be from 0.1 to 2 mM. We attribute the decrease of the sensitivity in comparison with turbidity assay to nonspecific binding of Rhod to FP. The decreased sensitivity in the high concentration range can be attributed to the readout method. While the turbidity assay intrinsically integrates the signal from full volume of hydrogel, fluorescence signal might only be read from a limited volume close to the gel surface, also dependent on gel density. Once this critical optical thickness of the gel is exceeded, the same signal will be measured for varying hydrogel coating thicknesses.

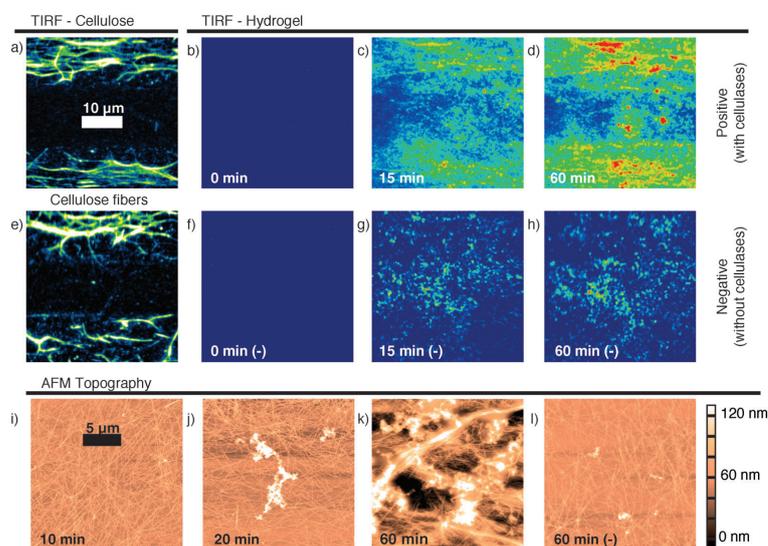
A small amount of nonspecific binding of Rhod to the solid substrates was observable, but in general was not problematic. Nonspecific binding is likely to be dependent on the type of substrate, its charge properties, and pretreatment conditions. Therefore, the performance of the HyReS system on a range of cellulose substrates was tested to determine its substrate compatibility profile. As shown in Figure 3c, the HyReS system with fluorescence detection was found to provide high signal-to-noise ratios on every substrate tested, including CMC, Avicel, Sigma  $\mu$ -crystalline cellulose powder, dilute acid pretreated hay, filter paper, and pretreated algal *Cladophora* cellulose. Nonspecific binding was not found to be a limitation, as indicated by the negative controls lacking the cellulolytic enzymes. The selectivity ratios of specific to non-specific signal ranged from 4.4 for Avicel to 751.9 for Sigma  $\mu$ -crystalline cellulose powder. All results were statistically significant using a one-sided t-test to  $P < 0.025$ . The system therefore has a wide applicability and seems to provide high signal-to-noise ratios on nearly any cellulose substrate susceptible to enzymatic degradation.

The pH-dependence of the assay was investigated by preparing the HyReS system at various pH values from 4.5 to 7.5 (see the Supporting Information). A pH of 5.0 or below was necessary for the reaction due to base catalyzed oxidation of Fe<sup>II</sup> to Fe<sup>III</sup> at higher pH values and consequent quenching of the reaction.<sup>[33]</sup> This low pH requirement might be limiting for this system for some applications as fungal cellulases have pH optima in the range of 4 to 6.5.<sup>[34]</sup> However, the HyReS system pH range (< pH 5) matches optimal conditions for many cellulolytic enzyme formulations (e.g., *Tr* and *A. niger* cocktails).<sup>[35,36]</sup>

Developing systems for real-time imaging of cellulose degradation is an important step towards improved enzyme formulations for biofuel development. In order to facilitate real-time imaging we used TIRF microscopy, which only samples molecules within an evanescent field extending away from the glass surface to a distance of a few hundred nanometers. This method restricts the excitation volume in a similar manner to confocal microscopy.<sup>[37]</sup> We were able to use nM quantities of the Rhod dye while simultaneously rejecting the fluorescent background and imaging the buildup of gel on the cellulose fibers. This setup eliminated the need to rinse away any unreacted Rhod before readout, significantly improving time resolution. The refractive index of the hydrogel is less than that of glass, therefore the critical angle requirement for TIRF was maintained and excitation light did not penetrate into the bulk even as the gel formed at the surface.

Figure 4 shows time-lapse TIRF imaging with the HyReS system. *Cladophora* cellulose was covalently labeled with a fluorescein derivative<sup>[38]</sup> (5-(4,6-dichlorotriazinyl) aminofluorescein, DTAF), and patterned in lines onto a cover slip (see Experimental Section). The sample was then imaged under liquid in the TIRF microscope. Under blue illumination (See "TIRF-cellulose", Figure 4a and e), patterned bands of labeled cellulose fibers were clearly visible at the top and bottom of the image, and reproduced the fibrous structure of the *Cladophora* cellulose in the TIRF image. The cellulose-free band forms the black stripe in the center of the image. Next, *Tr* enzymes and HyReS system including Rhod dye at 35 nM were added to the liquid, and images were collected over time under green illumination (Figure 4b–d). At time  $t=0$ , the gel had not yet formed and no Rhod signal was observable in the TIRF image (Figure 4B). By time  $t=60$  min., HyReS polymerization had incorporated Rhod into the hydrogel and the signal became observable in the TIRF image, mainly at locations where the cellulose was deposited, reproducing the substrate pattern with high fidelity (Figure 4d). This result indicated that reaction of the oligosaccharide hydrolysis products with the HyReS system components and initiation of polymerization occurred quickly enough to be localized to their site of production before the components could diffuse away from the fiber surface. Negative control experiments lacking the *Tr* enzyme mixtures (Figure 4e–h) showed only low non-specific signal that did not co-localize with the patterned substrate locations. The HyReS system therefore served as an imaging method and provided a fluorescent readout that increased from a low background to a high signal directly in response to hydrolysis of the substrate. To the best of our knowledge, such a localized chemical imaging system for cellulolytic activity has never been shown before using fluorescence detection. Such a method could provide distinct advantages in studies on cellulase synergy and susceptibility of cellulose substrates to degradation at specific locations (e.g., branch points, fibril ends, and/or crystalline faces).

To obtain more detailed information about the morphology of the hydrogel formation on solid substrates, we employed time resolved AFM imaging. DTAF-labeled cellulose was spin-coated uniformly onto a coverslip and the HyReS system was



**Figure 4.** Time-lapse TIRF(a–h) and AFM (i–l) imaging. Cellulose fibers were covalently labelled with a fluorescein derivative (DTAF) and patterned onto a cover slip. The stripes of patterned cellulose were clearly visible in blue TIRF illumination, while a middle band of the cover slip remained cellulose-free (a and e). The HyReS mixture including 35 nM Rhod and 2 mg mL<sup>-1</sup> *Tr* cellulases was added and sample was imaged under green illumination for 60 min (b, c, and d). Polymerization of the fluorescent hydrogel clearly co-localized with locations of micropatterned cellulose. The negative control experiment lacking *Tr* enzymes (images f, g, and h) showed only low non-specific background that did not co-localize with substrate locations. AFM height images (i–l) were obtained on cellulose that was deposited uniformly across the entire cover glass and exposed to the HyReS mixture. Panel (l) shows the negative control (60 min (-)) lacking *Tr* enzymes.

applied for varying amounts of time. Afterwards, samples were carefully rinsed and imaged in tapping mode in air (Figure 4i–l). The *Cladophora* cellulose formed a dense mat on the glass surface, consisting mostly of thin and long features corresponding to single cellulose fibers or small fiber bundles (Figure 4i). After 20 min, the HyReS system formed distinctive hydrogel features on the surface with heights of up to several hundred nm. The number and size of the features clearly increased with assay time. After an hour, large piles of hydrogel with heights of up to hundreds of nm and widths of several μm could be observed. This demonstrates the high signal amplification achieved by HyReS system because each hydroxyl radical initiates chain propagation that incorporates several hundred monomers into the growing gel. Additionally, the signal is integrated over time as the gel builds up. These amplification and integration mechanisms convert the glucose signal generated upon hydrolysis of nanometer-scale cellulose fibers into micrometer-scale hydrogel formations. At the same time, the size of the hydrogel formations originating from small cellulose features sets the intrinsic limit to the spatial resolution of presented method. The negative control showed small amounts of unspecific polymerization, consistent with our observations from TIRF imaging.

gration of the signal over time and space. We have furthermore presented results that demonstrate the HyReS system as an imaging platform for use in combination with TIRF microscopy and AFM, providing real-time imaging of cellulose hydrolysis with high spatial resolution. Our AFM imaging results demonstrate the extent of signal amplification that is possible when attempting to observe cellulose digestibility on nanometer-scale fibers. These unique features of the HyReS system can contribute to our understanding of how substrate structure affects enzymatic hydrolysis, and also move toward assaying the activity of individual cellulolytic complexes (i.e., cellulosomes) deposited onto cellulosic substrates. These results taken together establish the HyReS system as a competitive cellulase assay platform with the added advantage of spatially resolved localized chemical imaging.

## Experimental Section

**Materials:** Methacryloxyethyl thiocarbonyl rhodamine B (Rhod) was obtained from Polysciences Inc. (Warrington, PA, USA). Beta-1,4-endoglucanase from *T. emersonii* was purchased from Megazyme (Ireland). Glucose oxidase from *A. niger* and β-glucosidase from almonds were purchased from Sigma–Aldrich. All other reagents were obtained from Sigma–Aldrich and used without further purification. Composition of the standard reagent mixture

## Conclusion

We have shown that the HyReS system, comprising a mixture of cellulolytic enzymes, β-glucosidase, GOx, Fe<sup>II</sup>, ascorbic acid, PEG diacrylate, and Rhod is a versatile tool for detecting and imaging cellulolytic enzyme activity on a wide range of solid and soluble cellulose substrates. The system is compatible with turbidity detection on soluble substrates, and with fluorescence detection for insoluble substrates. Using the turbidity method, we have demonstrated glucose sensitivity in the low micromolar range which is on par with other bulk glucose determinations (e.g., DNS<sup>32</sup>). Analogous to conventional GOx/HRP systems, our system includes an amplification step as many vinylated monomers are incorporated into the growing gel for every hydroxyl radical initiator produced from glucose. Additionally, our system has other added advantages, such as localization of the signal to crystalline solid–liquid interfaces, and inte-

used for cellulase activity detection is shown in Table 1. All experiments used this standard mixture with slight variations noted in the text. Reagents were premixed prior to each experiment. Poly(propylene) 96-well were purchased from Grenier (Bio-One).

**Turbidity measurements on soluble substrates:** For the glucose calibration plot (Figure 2a), wells of a 96-well plate were filled with 100  $\mu\text{L}$  of acetate buffer containing twice the target concentration of the HyReS system (Table 1). An equal volume of acetate buffer (100  $\mu\text{L}$ ) containing twice the target glucose concentration was added. Monitoring of the absorbance (600 nm) began immediately and continued for 30 min inside a plate reader (Tecan M1000 Pro) at 37 °C. The endoglucanase assay (Figure 2b) was performed similarly, using CMC in place of glucose. CMC (degree of substitution: 0.60–0.95) was dissolved in acetate buffer, pH 4.5. Each well was filled with a total volume of 100  $\mu\text{L}$  containing the indicated amount of CMC, cellulolytic enzymes, and the standard HyReS reagent mixture (without Rhod dye). The plate was incubated at 37 °C inside a plate reader and absorbance was measured continuously at 550 nm. The reported errors correspond to the standard deviation of at least three independent measurements.

**Fluorescence measurements on solid substrates:** Filter paper (Whatman #1, FP) was cut into disks (6 mm diameter, 2.5 mg cellulose), placed into the wells of a 96-well plate and used as the cellulose substrate. For calibration of the assay, 5  $\mu\text{L}$  of glucose standards were allowed to soak into the FP disks, followed by addition of 195  $\mu\text{L}$  of HyReS system (lacking cellulases). After incubation at 37 °C, unreacted monomer was removed with a water rinse using a microplate strip washer (ELx50, BioTek). Fluorescence at 580 nm was measured in a plate reader with excitation at 555 nm. For the cellulase assays, a total liquid volume of 200  $\mu\text{L}$  containing cellulolytic enzymes (range 0–1  $\text{mg mL}^{-1}$ ) together with 3.5  $\mu\text{M}$  Rhod and the standard reagent mixture (Table 1) was added to each well. After incubation at 37 °C, polymerization was stopped by removing unreacted monomer with a water rinse using ELx50 Microplate Strip Washer (BioTech). Fluorescence was measured immediately with a plate reader (M1000pro, Tecan) with excitation at 555 nm, and emission at 580 nm. The reported errors correspond to the standard deviation of at least three independent measurements.

**DTAF-grafted cellulose fibers (DTAF-CF):** Cellulose fibers were extracted from fresh *Cladophora* algae according to published protocols.<sup>[39,40]</sup> Noncellulosic cell components were first extracted in 98% ethanol at 50 °C for 24 h. Solid material was filtered and subsequently boiled for 2 h in 0.1 M NaOH. After exchanging the NaOH solution, cellulose was again extracted at 80 °C overnight. Afterwards, the sample was immersed in 0.05 M HCl at room temperature for 12 h, filtered, thoroughly washed with water and freeze-dried. In order to obtain cellulose microcrystals, the sample was further acid hydrolyzed in 40%  $\text{H}_2\text{SO}_4$  at 70 °C for 12 h. After extensive centrifugal separation and washing, cellulose was dialyzed against deionized water and the suspension was stored in water at 4 °C in darkness for up to several weeks prior to use.

*Cladophora* cellulose fibers obtained in this way were covalently labeled with the fluorescent dye DTAF according to previously published protocols.<sup>[38,42]</sup> In short, 5 mg of DTAF was dissolved in 1 mL of 0.2 M NaOH. The resulting solution was mixed with 500  $\mu\text{L}$  of the cellulose suspension in water and reacted for 24 h at room temperature. Unreacted dye was removed by centrifugal washing five times followed by dialysis against water.

**Cellulose micropatterning:** Round cover slips (borosilicate, 22 mm dia., 0.2 mm thickness, Thermo Fisher) were aminosilanized follow-

ing previously published procedures.<sup>[41]</sup> DTAF-labeled cellulose fibers were patterned on aminosilanized cover slips under flow in a PDMS microfluidic channel. A PDMS mold with two parallel channels 100  $\mu\text{m}$  wide, 28  $\mu\text{m}$  high and 2 cm long, spaced 15  $\mu\text{m}$  apart was produced using standard soft lithography methods, and applied onto an aminosilanized glass surface and cured at 65 °C overnight. A suspension of DTAF-CF was sonicated for 3 min to disperse fibrils, introduced into the channels and incubated for 5 min. The negatively charged DTAF-CFs adhered to positively charged aminosilanized glass surface. Afterwards, the channels were flushed with water to remove weakly bound fibers. The flow channel was then removed, and surfaces were blocked for 2 h by exposure to a solution of 2  $\text{mg mL}^{-1}$  BSA in acetate buffer (20 mM, pH 4.5) followed by rinsing with water.

**Total internal reflection fluorescence microscopy:** Fluorescence imaging was carried out in TIRF excitation on a custom-built multi-color TIRF microscope, similar to the instrument described previously by Gump et al.<sup>[43]</sup> Blue DTAF dye was excited by the 488 nm line and Rhod by the 561 nm line of the TOPTICA iChrome MLE-LFA laser through a 100 $\times$ , NA 1.49 oil immersion objective lens (Nikon Apochromat). We used ET525/36 and HC600/37 emission filters mounted in Optosplit III (Carin Research) for detection of DTAF and Rhod fluorescence, respectively. The emitted light was detected using a 1024 $\times$ 1024 pixel back-illuminated EMCCD camera (Andor iXon3 888).

The cover glass with micropatterned lines of DTAF-CFs was placed in a liquid-tight holder and mounted on the fluorescence microscope. First, cellulose fibers in buffer were imaged under buffer to verify patterning fidelity. To visualize hydrogel build-up in real time, 300  $\mu\text{L}$  of the standard reagent mixture supplemented with 2  $\text{mg mL}^{-1}$  *Tr* cellulolytic enzymes, and 35 nM Rhod were added onto the sample. Time series were recorded in blue and green channels with an integration time of 0.5 s per frame and 10 s between acquisitions. The Peltier-cooled CCD chip was typically operated at a temperature of  $-80$  °C and an electron multiplication gain of 150 $\times$  and 200 $\times$  was used for blue and green channels respectively.

**Atomic force microscopy:** Measurements were carried out using MFP-3D AFM (Asylum Research) in combination with AC160 cantilevers (resonance frequency: 300 kHz, spring constant: 27  $\text{N m}^{-1}$ , Olympus). All imaging studies were done in tapping mode with amplitude of  $\sim 100$  nm. DTAF-CFs were spin coated onto an aminosilanized cover slip (3000 rpm, 60 s). The standard hydrogel reagent mix including 1  $\text{mg mL}^{-1}$  *Tr* cellulases was added to the cover slip and sample was incubated for varying amounts of time at 37 °C. Polymerization was stopped by a gentle rinse in a beaker of ultrapure water. The sample was blow dried with gentle nitrogen stream and imaged in air.

## Acknowledgements

The authors gratefully acknowledge funding from the European Research Council and the Nanosystems Initiative Munich. M.A.N. acknowledges funding from Society in Science–The Branco Weiss Fellowship administered by the Swiss Federal Institute of Technology (ETH Zürich).

**Keywords:** biomass · cellulase enzymes · radical reactions · TIRF imaging · *Trichoderma reesei*

- [1] M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, T. D. Foust, *Science* **2007**, *315*, 804–807.
- [2] D. Klein-Marcuschamer, P. Oleskowicz-Popiel, B. A. Simmons, H. W. Blanch, *Biotechnol. Bioeng.* **2012**, *109*, 1083–1087.
- [3] P. Kumar, D. M. Barrett, M. J. Delwiche, P. Stroeve, *Ind. Eng. Chem. Res.* **2009**, *48*, 3713–3729.
- [4] R. Peterson, H. Nevalainen, *Microbiology* **2012**, *158*, 58–68.
- [5] W. Helbert, H. Chanzy, T. L. Husum, M. Schülein, S. Ernst, *Biomacromolecules* **2003**, *4*, 481–487.
- [6] Y. H. P. Zhang, J. Hong, X. Ye in *Methods in Molecular Biology*, Humana Press, Totowa, NJ, **2009**, pp. 213–231.
- [7] A. Chandrasekaran, R. Bharadwaj, J. I. Park, R. Sapra, P. D. Adams, A. K. Singh, *J. Proteome Res.* **2010**, *9*, 5677–5683.
- [8] M. Dashtban, M. Maki, K. T. Leung, C. Mao, W. Qin, *Crit. Rev. Biotechnol.* **2010**, *30*, 302–309.
- [9] D. R. Ivanen, N. L. Rongjina, S. M. Shishlyannikov, G. I. Litviakova, L. S. Isaeva-Ivanova, K. A. Shabalina, A. A. Kulminskaya, *J. Microbiol. Methods* **2009**, *76*, 295–300.
- [10] D. J. Coleman, M. J. Studler, J. J. Naleway, *Anal. Biochem.* **2007**, *371*, 146–153.
- [11] Y. H. Percival Zhang, M. E. Himmel, J. R. Mielenz, *Biotechnol. Adv.* **2006**, *24*, 452–481.
- [12] Z. Xiao, R. Storms, A. Tsang, *Biotechnol. Bioeng.* **2004**, *88*, 832–837.
- [13] R. Mullings, *Enzyme Microb. Technol.* **1985**, *7*, 586–591.
- [14] S. Kongruang, M. K. Bothwell, J. McGuire, M. Zhou, R. P. Haugland, *Enzyme Microb. Technol.* **2003**, *32*, 539–545.
- [15] V. Harjunpää, J. Helin, A. Koivula, M. Siika-aho, T. Drakenberg, *FEBS Lett.* **1999**, *443*, 149–153.
- [16] N. Cruys-Bagger, S. F. Badino, R. Tokin, M. Gontsarik, S. Fathalinejad, K. Jensen, M. D. Toscano, T. H. Sørensen, K. Borch, H. Tatsumi, P. Våljamäe, *Enzyme Microb. Technol.* **2014**, *58*–59, 68–74.
- [17] P. Bubner, J. Dohr, H. Plank, C. Mayrhofer, B. Nidetzky, *J. Biol. Chem.* **2012**, *287*, 2759–2765.
- [18] T. Ganner, P. Bubner, M. Eibinger, C. Mayrhofer, H. Plank, B. Nidetzky, *J. Biol. Chem.* **2012**, *287*, 43215–43222.
- [19] B. G. Saar, Y. Zeng, C. W. Freudiger, Y.-S. Liu, M. E. Himmel, X. S. Xie, S.-Y. Ding, *Angew. Chem. Int. Ed.* **2010**, *49*, 5476–5479; *Angew. Chem.* **2010**, *122*, 5608–5611.
- [20] J. M. Fox, P. Jess, R. B. Jambusaria, G. M. Moo, J. Liphardt, D. S. Clark, H. W. Blanch, *Nat. Chem. Biol.* **2013**, *9*, 356–361.
- [21] S. Y. Ding, Y. S. Liu, Y. Zeng, M. E. Himmel, J. O. Baker, E. A. Bayer, *Science* **2012**, *338*, 1055–1060.
- [22] D. J. Dagel, Y.-S. Liu, L. Zhong, Y. Luo, M. E. Himmel, Q. Xu, Y. Zeng, S.-Y. Ding, S. Smith, *J. Phys. Chem. B* **2011**, *115*, 635–641.
- [23] E. Neyens, J. Baeyens, *J. Hazard. Mater.* **2003**, *98*, 33–50.
- [24] F. S. Dainton, P. H. Seaman, *J. Polym. Sci.* **1959**, *39*, 279–297.
- [25] A. S. Sarac, *Prog. Polym. Sci.* **1999**, *24*, 1149–1204.
- [26] R. Shenoy, M. W. Tibbitt, K. S. Anseth, C. N. Bowman, *Chem. Mater.* **2013**, *25*, 761–767.
- [27] L. M. Johnson, C. A. DeForest, A. Pendurti, K. S. Anseth, C. N. Bowman, *ACS Appl. Mater. Interfaces* **2010**, *2*, 1963–1972.
- [28] B. J. Berron, L. M. Johnson, X. Ba, J. D. McCall, N. J. Alvey, K. S. Anseth, C. N. Bowman, *Biotechnol. Bioeng.* **2011**, *108*, 1521–1528.
- [29] H. D. Sikes, R. R. Hansen, L. M. Johnson, R. Jenison, J. W. Birks, K. L. Rowlen, C. N. Bowman, *Nat. Mater.* **2008**, *7*, 52–56.
- [30] G. Gefen, M. Anbar, E. Morag, R. Lamed, E. A. Bayer, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10298–10303.
- [31] C. Gonçalves, R. M. Rodríguez-Jasso, N. Gomes, J. A. Teixeira, I. Belo, *Anal. Methods* **2010**, *2*, 2046.
- [32] B. Halliwell, J. Gutteridge, *FEBS Lett.* **1992**, *307*, 108–112.
- [33] T. Tabakova, D. Andreeva, A. Andreev, C. Vladov, I. Mitov, *J. Mater. Sci. Mater. Electron.* **1992**, *3*, 201–205.
- [34] L. N. Anderson, D. E. Culley, B. A. Hofstad, L. M. Chauvigne-Hines, E. M. Zink, S. O. Purvine, R. D. Smith, S. J. Callister, J. M. Magnuson, A. T. Wright, *Mol. Biosyst.* **2013**, *9*, 2992–3000.
- [35] T. Wang, X. Liu, Q. Yu, X. Zhang, Y. Qu, P. Gao, T. Wang, *Biomol. Eng.* **2005**, *22*, 89–94.
- [36] P. L. Hurst, J. Nielsen, P. A. Sullivan, M. G. Shepherd, *Biochem. J.* **1977**, *165*, 33–41.
- [37] J. S. Luterbacher, L. P. Walker, J. M. Moran-Mirabal, *Biotechnol. Bioeng.* **2013**, *110*, 108–117.
- [38] M. Santa-Maria, T. Jeoh, *Biomacromolecules* **2010**, *11*, 2000–2007.
- [39] T. Imai, J. Sugiyama, *Macromolecules* **1998**, *31*, 6275–6279.
- [40] T. Imai, J.-L. Putaux, J. Sugiyama, *Polymer* **2003**, *44*, 1871–1879.
- [41] M. A. Jobst, C. Schoeler, K. Malinowska, M. A. Nash, *J. Vis. Exp.* **2013**, *82*, e50950.
- [42] T. Abitbol, A. Palermo, J. M. Moran-Mirabal, E. D. Cranston, *Biomacromolecules* **2013**, *14*, 3278–3284.
- [43] H. Gump, S. W. Stahl, M. Strackharn, E. M. Puchner, H. E. Gaub, *Rev. Sci. Instrum.* **2009**, *80*, 063704.

Received: May 16, 2014  
Revised: June 12, 2014  
Published online on August 12, 2014

## Supporting Information

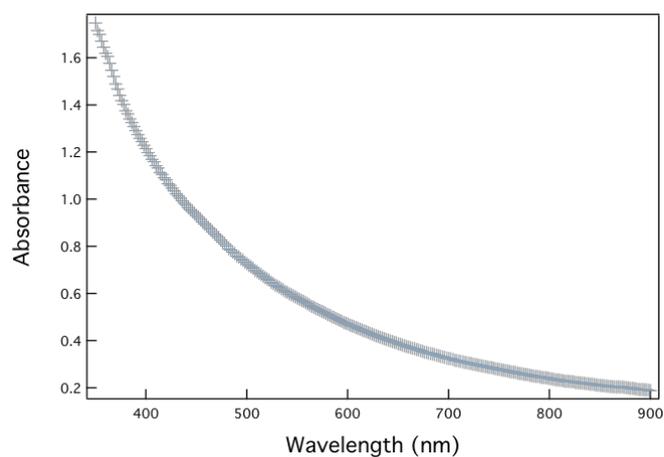
© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2014

### **Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity**

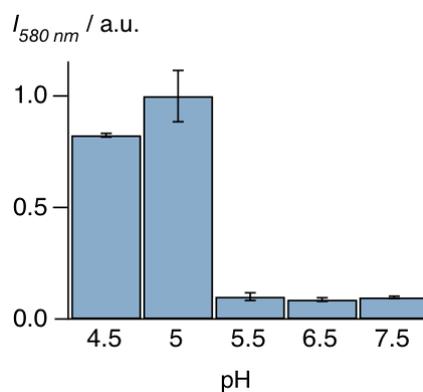
Klara H. Malinowska, Tobias Verdorfer, Aylin Meinhold, Lukas F. Milles, Victor Funk, Hermann E. Gaub, and Michael A. Nash<sup>\*[a]</sup>

cssc\_201402428\_sm\_miscellaneous\_information.pdf

## Supporting figures



**Supporting Figure 1.** Absorbance spectrum of the polymerized hydrogel. 20 mM CMC were mixed with 1mg/ml T. reesei enzymes and the hydrogel standard mix. After the full polymerization absorbance was measured using a plate reader (M1000 pro, Tecan).



**Supporting Figure 2.** pH-dependence of the HyReS signal on filter paper using fluorescence detection. Base catalyzed oxidation of the Fe(II) catalyst quenches the reaction above pH 5.





### 3.6 PUBLICATION P6: Quantifying Synergy, Thermostability, and Targeting of Cellulolytic Enzymes and Cellulosomes with Polymerization-Based Amplification

The following publication develops the cellulase activity assay introduced in section publication P5 further to allow for a label-free, parallelized and highly sensitive determination of cellulolytic efficiency on solid substrates. This one-pot implementation relies on monitoring the attenuation of cellulose autofluorescence as an opaque hydrogel polymerizes on top of the sample in direct response to glucose produced during saccharification. It enabled a differentiation of cellulolytic enzyme performance based on thermostability, substrate targeting, and synergistic effects.

# Quantifying Synergy, Thermostability, and Targeting of Cellulolytic Enzymes and Cellulosomes with Polymerization-Based Amplification

Klara H. Malinowska, Thomas Rind, Tobias Verdorfer,  
Hermann E. Gaub and Michael A. Nash

*published in*

Analytical Chemistry, 87(14), 7133 – 7140, (2015)

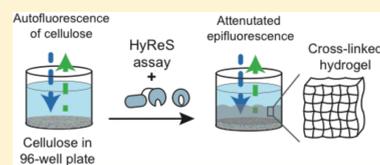
*Reprinted from [118], with permission from American Chemical Society  
Copyright 2015, American Chemical Society*

## Quantifying Synergy, Thermostability, and Targeting of Cellulolytic Enzymes and Cellulosomes with Polymerization-Based Amplification

Klara H. Malinowska, Thomas Rind, Tobias Verdorfer, Hermann E. Gaub, and Michael A. Nash\*

Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

**ABSTRACT:** We present a polymerization-based assay for determining the potency of cellulolytic enzyme formulations on pretreated biomass substrates. Our system relies on monitoring the autofluorescence of cellulose and measuring the attenuation of this fluorescent signal as a hydrogel consisting of poly(ethylene glycol) (PEG) polymerizes on top of the cellulose in response to glucose produced during saccharification. The one-pot method we present is label-free, rapid, highly sensitive, and requires only a single pipetting step. Using model enzyme formulations derived from *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Talaromyces emersonii* and recombinant bacterial minicellulosomes from *Clostridium thermocellum*, we demonstrate the ability to differentiate enzyme performance based on differences in thermostability, cellulose-binding domain targeting, and endo/exoglucanase synergy. On the basis of its ease of use, we expect this cellulase assay platform to be applicable to enzyme screening for improved bioconversion of lignocellulosic biomass.



A long-standing goal in the chemical sciences has been to develop biobased systems for efficient conversion of naturally occurring plant cell wall biomass into soluble sugars. This second-generation route toward renewable fuels and chemicals has the potential to alter the international landscape governing energy and chemical commodity markets in the near future. Efficient production of soluble fermentable sugars from lignocellulosic biomass would provide a valuable input into standard fermentation processes, or alternatively feed into processes involving synthetic microorganisms for the production of a wide range of chemicals, pharmaceuticals, and other valuable products.

In order to improve biological enzyme-based conversion systems for saccharification of lignocellulosic biomass, enzyme formulations are being steadily improved through a combination of directed evolution and semirational design strategies.<sup>1–3</sup> In terms of screening for enzyme activity, lignocellulosic bioconversion systems present a unique challenge.<sup>4</sup> The lignocellulosic substrates are not easily standardized, and the mass content of the primary components (lignin, hemicelluloses, and cellulose) will vary widely depending on the nature of the feedstock, where it was grown, and how it was pretreated.<sup>5</sup> Also, microscopic structure of a substrate plays a key role in enzyme adsorption, kinetics, and efficiency, as shown by recent spatially and time-resolved studies utilizing fluorescence<sup>6–9</sup> and atomic force microscopy.<sup>10–12</sup> New assays for evaluating the effectiveness of enzyme formulations on real-world industrially relevant pretreated biomass that are straightforward to implement, compatible on natural substrates, rapid, and highly sensitive are therefore clearly needed.

Here we present the use of a label-free hydrogel reagent signaling system (HyReS) for assaying hydrolysis of lignocellulosic biomass. Formation of a cross-linked hydrogel at the location of glucose production attenuates the autofluorescence of cellulose and is used for quantifying total cellulolytic activity.

The HyReS assay has an ability to rapidly quantify activity, thermostability, exo/endo synergy, and targeting effects in cellulolytic enzyme formulations as well as to show digestibility variations between different industrially relevant types of biomass.<sup>13</sup>

**Assay Principle.** Most of the commonly used cellulase activity assays rely on absorption or fluorescent dyes for signal detection. Those include the IUPAC-standardized colorimetric filter paper assay (FPA) in traditional<sup>14,15</sup> and microplate<sup>16–18</sup> formats, as well as bioenzymatic assays such as glucose oxidase (GOx)/horseradish peroxidase systems with fluorescence detection<sup>19,20</sup> and hexokinase/glucose-6-phosphate dehydrogenase systems based on nicotinamide adenine dinucleotide absorbance in the near-UV.<sup>21,22</sup> Novel glucose detection techniques also use fluorescent dyes for readout.<sup>20,23–25</sup> However, both cellulose and lignin exhibit autofluorescence,<sup>26,27</sup> a property that was previously used to map changes in cellulose and lignin content and their spatial distribution during biomass pretreatment<sup>28</sup> and to track changes in biomass structure along with localization of cellulolytic enzymes in real time.<sup>6</sup> As we show in this work, the intrinsic fluorescence of biomass can also be exploited to eliminate the need for dyes and labels in cellulolytic assays altogether.

Activity of cellulolytic cocktails is routinely assayed on standardized substrates such as filter paper, carboxymethyl cellulose (CMC), or Avicel<sup>29</sup> which, though readily available and easy to handle, have properties distinctly different from those of industrially relevant pretreated biomass.<sup>30</sup> The need for employing real lignocellulosic substrates in screening of cellulases is recognized in the community.<sup>4,31,32</sup> Several solutions have been proposed including the use of finely

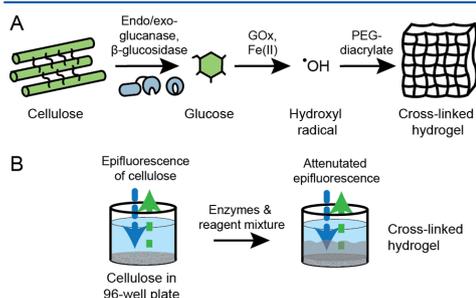
Received: March 10, 2015

Accepted: June 12, 2015

Published: June 26, 2015

ground substrate in suspension<sup>33</sup> and preparation of substrate discs from biomass sheets.<sup>34</sup>

The principle of our label-free HyReS system is the attenuation of lignocellulose autofluorescence due to light scattering on a hydrogel film formed at the location of glucose production (Figure 1). The GOx/Fe(II) system, described



**Figure 1.** Schematic overview of the label-free HyReS system. (A) Cellulolytic enzymes (e.g., endo/exoglucanase and  $\beta$ -glucosidase) hydrolyze lignocellulosic biomass producing glucose. Saccharification products are oxidized by GOx creating hydrogen peroxide that reacts with an  $\text{Fe}^{2+}$  Fenton reagent to produce short-lived hydroxyl radicals. The hydroxyl radicals initiate free radical polymerization of a PEG hydrogel, cross-linking PEG at the surface of the cellulosic substrate. (B) Autofluorescence of cellulose in the near-UV range is used to detect the hydrogel film. Prior to hydrogel formation, the optical path between the excitation source and detector remains unobstructed and the epifluorescence signal is collected. Once glucose release initiates gel formation, both excitation and emission light is scattered by the turbid gel, resulting in signal attenuation.

previously in detail by our group and others,<sup>25,35,36</sup> enables selective polymerization of poly(ethylene glycol) (PEG) hydrogel in the presence of glucose. Glucose is oxidized by GOx, and the resulting hydrogen peroxide further reacts with a Fenton reagent ( $\text{Fe}^{2+}$  ions) producing  $\cdot\text{OH}$  radicals. The resultant hydroxyl radicals then initiate free radical polymerization of PEG diacrylate, resulting in a densely cross-linked gel. Radical polymerization serves as a signal amplification step since multiple monomers are incorporated into the hydrogel network for each released glucose molecule. The Fenton reagent can then be regenerated in the reaction of  $\text{Fe}^{3+}$  with ascorbic acid.<sup>37</sup> Substrate autofluorescence is measured in epillumination mode from above, and formation of turbid gel is detected via fluorescence signal attenuation.

## RESULTS AND DISCUSSION

**Substrate Characterization.** We prepared 6 mm discs of pretreated biomass (napier grass and miscanthus, Figure 2, parts A and B) using standard laboratory equipment. As opposed to filter paper, pretreated biomass contains traces of lignin which influences its digestibility. Figure 2C shows Raman spectra of substrates with bands attributed to cellulose (380, 435, 1095, 1123, 1377, and 2985  $\text{cm}^{-1}$ ) present in all samples and lignin (1600  $\text{cm}^{-1}$ ) absent in filter paper.<sup>38</sup> Autofluorescence spectral scans of all substrates (Figure 2D–F) were dominated by broad cellulose peaks with maxima at  $\lambda_{\text{ex}} = 365$  nm and  $\lambda_{\text{em}} = 430$  nm.<sup>27</sup> An additional broad shoulder at longer wavelengths present in the spectra of napier grass and miscanthus originates from lignin,<sup>28</sup> while multiple bands at

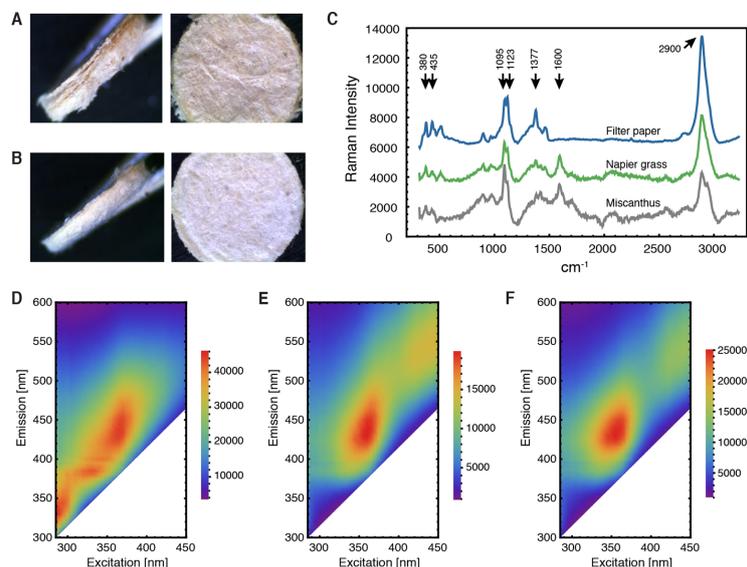
shorter wavelengths in the spectrum of filter paper were attributed to optical brighteners.<sup>39</sup> These results identify 365/430 nm wavelength as an optimal choice for universal detection of biomass substrates using the HyReS assay.

We note that drying of biomass can affect the crystalline structure and digestibility. The polymerization assay, however, is also compatible with never-dried biomass. In our experience, simple centrifugation of a biomass slurry in a 96-well plate results in a compacted cellulose sediment at the bottom of the wells that is sufficiently cohesive to withstand gentle addition of liquid, allowing for the same autofluorescence measurement (described below) to be performed with never-dried biomass.

**Assay Validation.** In a proof-of-principle experiment, we used the label-free HyReS system to quantify cellulolytic activity of a *Trichoderma reesei* enzyme cocktail. Cellulases over a concentration range from 0 to 100  $\mu\text{g mL}^{-1}$  were premixed with components of the label-free HyReS assay and preheated to 37 °C. Discs of pretreated biomass were placed in wells of a microtiter plate, and the assay mix was added. The plate was incubated at 37 °C, and cellulose fluorescence was monitored over time.

The resulting time-resolved autofluorescence attenuation patterns were similar for both biomass samples (Figure 3, parts A and B, top). During the first 20 min of incubation, fluorescence intensity decreased until a plateau was reached at approximately 80% of initial signal intensity. This behavior was consistent for all wells including the negative control without cellulolytic enzymes present. This initial decrease was due to changes in the liquid meniscus shape at early time points, confirmed by time-lapse video microscopy of the wells from the side. Control measurements indicated no significant photobleaching of the sample under the experimental conditions. After this initial decline in fluorescence, a second drop in signal intensity down to approximately 40% of the initial fluorescence was observed. The second drop in autofluorescence was the result of formation of a thin, opaque hydrogel film on the substrate surface. Afterward, the fluorescence intensity rose slightly until the end of the measurement, which can be explained by a gradual evaporation of liquid from the wells, resulting in a decreased path length through the liquid.

The time at which the hydrogel film formed and attenuated the fluorescence signal was dependent on the concentration of cellulases present in the sample. Higher concentrations of cellulolytic enzymes resulted in a faster rise of glucose concentration in the vicinity of the substrate and led to earlier formation of the hydrogel film. To quantify this effect, we developed a data analysis method involving normalization, smoothing, and numerical differentiation of fluorescence time traces (see the Experimental Section). The maximum value of the derivative corresponds to the fastest signal attenuation per unit time and, consequently, to the most rapid rate of hydrogel production (Figure 3, parts A and B, bottom). The time at which the maximal signal change occurred plotted against the concentration of cellulolytic enzymes on a log scale (Figure 3C) shows that the relation between cellulose concentration and attenuation time is nonlinear. The assay is sensitive down to 3 and 1  $\mu\text{g mL}^{-1}$  *T. reesei* enzymatic cocktail within 200 min on napier grass and miscanthus, respectively. Longer incubation times can increase sensitivity even further. In terms of absolute glucose sensitivity, our prior work described calibration of a similar HyReS assay that did not rely on substrate



**Figure 2.** Pretreated biomass substrate characterization. Side and top view of cylindrical discs (6 mm in diameter) produced from (A) napier and (B) miscanthus perennial grass. (C) Raman spectra of pretreated biomass substrates using 568 nm excitation. Bands at 380, 435, 1095, 1123, 1377, and 2985  $\text{cm}^{-1}$  were attributed to cellulose, with lignin contribution visible at 1600  $\text{cm}^{-1}$ . Spectra were background-corrected and vertically offset for clarity. Excitation/emission autofluorescence spectral scans of (D) filter paper, (E) miscanthus, and (F) napier grass exhibited a prominent cellulose peak at  $\approx 365/430$  nm  $\lambda_{\text{ex}}/\lambda_{\text{em}}$ . A lignin shoulder at longer wavelengths was present in miscanthus and napier grass samples.

autofluorescence. For that system, linear dynamic range was between 0.05 and 5 mM glucose.<sup>25</sup>

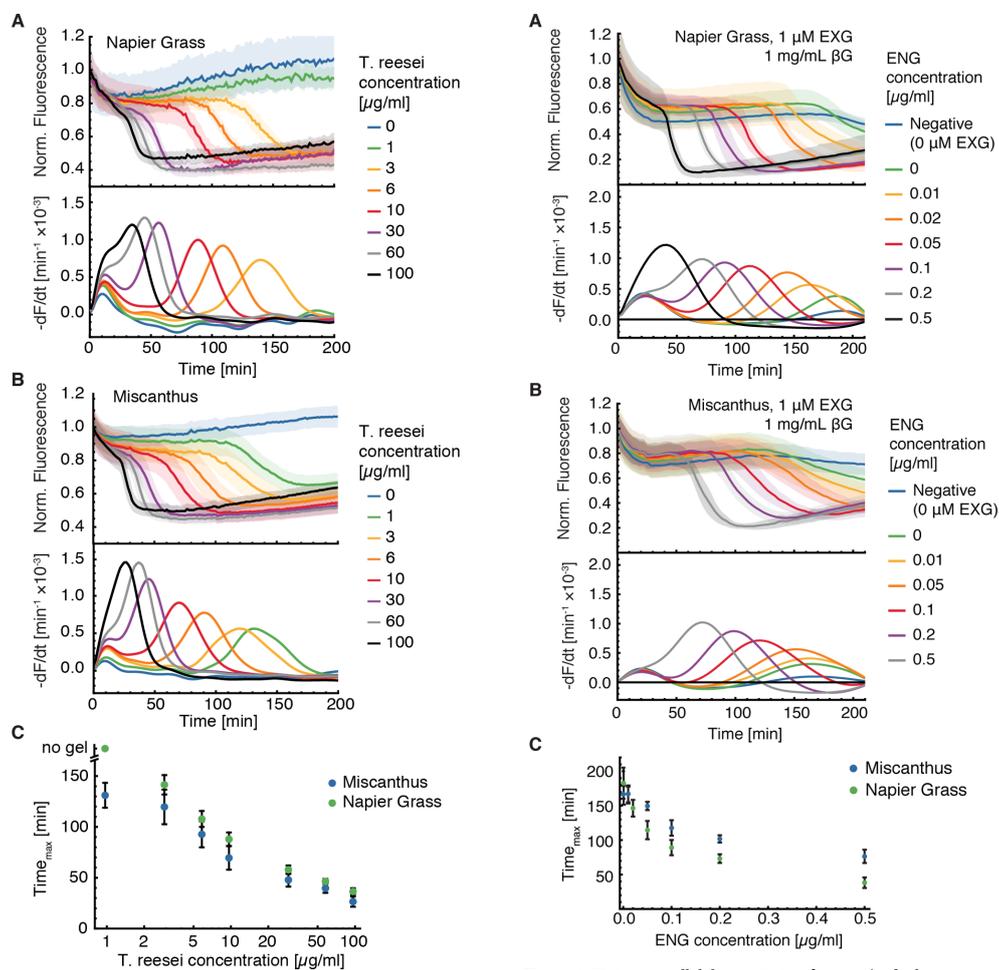
In an analogous experiment we tested the ability of the system to detect differences in combined cellulolytic activity of exoglucanase (cellobiohydrolase I, EXG), endoglucanase (ENG), and  $\beta$ -glucosidase ( $\beta$ G) upon changes in ENG concentration. The concentrations of EXG and  $\beta$ G were kept constant at 1  $\mu\text{M}$  and 1  $\text{mg mL}^{-1}$ , respectively, while the concentration of ENG was varied between 0 and 0.5  $\mu\text{M}$ . The position of the maximum rate of change of the fluorescence signal correlated well with enzymatic activity (Figure 4). Mixtures containing more ENG produced glucose faster and thus enabled the formation of a hydrogel film in a much shorter time.

**Quantifying Synergistic and Targeting Effects.** Synergy, or an enhanced activity of different types of cellulases acting together, is an important design parameter for development of multienzyme formulations.<sup>40,41</sup> However, synergistic effects in complex mixtures of enzymes can be hard to predict, and the extent of synergy is strongly substrate-dependent, competition being the most extreme case.<sup>42</sup> Also, the efficiency of targeting enzymes to the substrate by cellulose binding modules (CBMs) is strongly dependent on the microscopic structure of biomass.<sup>43</sup> Because of these complex enzyme–enzyme and enzyme–substrate dependencies, it is important to experimentally evaluate various cellulase compositions on relevant biomass sources to adequately judge synergy and targeting effects.

To address this point, we used the label-free HyReS assay to assess cellulolytic activity of an enzyme mixture containing 1  $\mu\text{M}$  EXG, 0.1  $\mu\text{M}$  ENG, and 1  $\text{mg mL}^{-1}$   $\beta$ G on miscanthus

and napier grass (Figure 5). While EXG alone and combined with  $\beta$ G was equally effective on both substrates, the rate of glucose production from napier grass by ENG (with and without  $\beta$ G) was much higher than from miscanthus. As expected, combining EXG and ENG led to drastically increased hydrolysis rates on both substrates. For example, the  $T_{\text{max}}$  values for individual EXG and ENG on miscanthus were 109 and 127 min, respectively, which corresponds to the activity of approximately 4 and 1  $\text{mg mL}^{-1}$  of *T. reesei* enzymatic mixture. The combined EXG/ENG mixture had  $T_{\text{max}}$  of 61 min, which compares with the hydrolytic potential of approximately 15  $\text{mg mL}^{-1}$  of *T. reesei* cellulases. The activity of the EXG/ENG mixture was much higher than the sum of activities of the separate EXG and ENG enzymes independently, therefore indicating their synergistic action on solid cellulose. It is worth noting that a EXG/ENG/ $\beta$ G formulation was more effective on pretreated napier grass than on miscanthus, contrary to the *T. reesei* cocktail which hydrolyzed the latter substrate preferably (Figure 3).

CBMs are known to increase cellulolytic activity both when connected to single catalytic domains by flexible linkers and when incorporated into cellulosomal scaffolding.<sup>44,45</sup> We evaluated the effect of CBM incorporation of cellulose decomposition by comparing trimodular Cel8A-loaded minicellulosomes with and without a CBM in the scaffold. Concentrations of 0.2  $\mu\text{M}$  of minicellulosomes (corresponding to 0.6  $\mu\text{M}$  of endoglucanase) showed a significant increase in hydrolysis rate on various biomass types when loaded onto a miniscaffold containing a CBM domain (Figure 6). This was due to the high affinity of CBM to cellulose that prolonged the bound lifetime of the catalytic domains on the substrate and



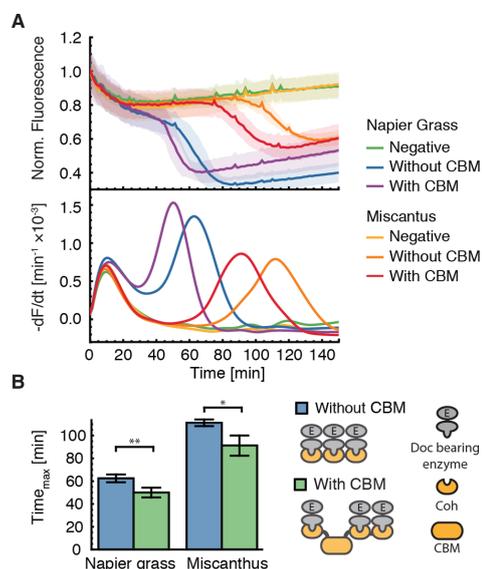
**Figure 3.** Detecting the cellulolytic activity of the *T. reesei* enzymatic cocktail. Attenuation of (A) napier grass and (B) miscanthus autofluorescence by the hydrogel film formed in response to enzymatic glucose production. (A and B, top) Changes of fluorescence signal in time. The shadowed area represents standard deviation of five independent measurements. (A and B, bottom) First derivative of fluorescence signal over time. (C) Time at which the peak in fluorescence derivative occurs plotted against the *T. reesei* enzymatic cocktail concentration. Lower  $T_{\max}$  values represent high enzymatic activity.

increased their concentration in the immediate proximity of the substrate.

**Quantifying Thermostability of Cellulases.** One more application that we foresee for the HyReS assay is selecting cellulases for thermostability, a quality which can increase their lifetime under the harsh conditions required for bioprocessing.<sup>46</sup> As an example, two cellulases, EXG and ENG, were heat-shocked at 80 °C for variable time intervals from 0 to 90 min.

**Figure 4.** Detecting cellulolytic activity of an exo/endoglucanase mix by measuring attenuation of (A) napier grass and (B) miscanthus autofluorescence. (A and B, top) Changes of epifluorescent signal vs time. Shadowed areas represent the standard deviation of five measurements. (A and B, bottom) First derivative of fluorescence signal vs time. (C) Time at which the peak in fluorescence derivative occurs plotted against the ENG concentration. The concentration of EXG was kept constant at 1 µM.

Afterward, their activity on filter paper was evaluated using the dye-free HyReS assay (Figure 7). The gel formation in presence of ENG was fast regardless of prolonged heat exposure, indicating that activity of this thermophilic enzyme remained largely unaffected by temperature. On the contrary, activity of the EXG decreased drastically after 5 min of heat shock, and after 9 min gel formation was not detectable, indicating total loss of activity of this mesophilic enzyme.

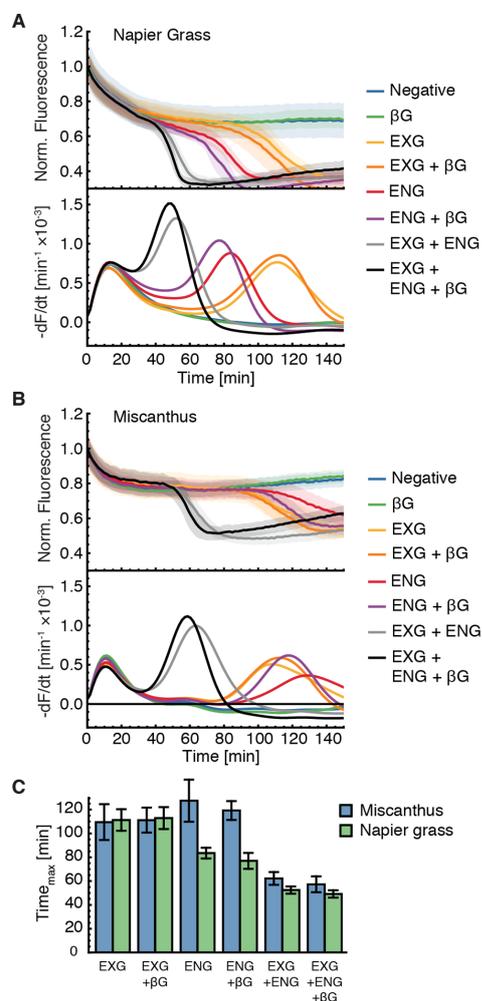


**Figure 5.** Activity of trimodular endoglucanase-loaded minicellulomes on pretreated napier and miscanthus grasses. (A, top) Changes of epifluorescence signal in time. Shadow area represents standard deviation of five measurements. (A, bottom) First derivative of fluorescence signal over time. (B) Time at which the peak in fluorescence derivative occurs for miniscrafts with and without CBM (see inset). \*  $P < 0.01$ , \*\*  $P < 0.005$  in two-tailed unpaired Student  $t$  test.

## CONCLUSIONS

Several qualities significantly differentiate the label-free HyReS system from other cellulolytic activity assays, and from our prior work.<sup>25</sup> The simplicity of preparation of substrate discs from virtually any type of pretreated biomass allows the assessment of hydrolytic potential of enzymatic cocktails in conditions relevant to the biomass-to-bioenergy industry. This feature circumvents the issue of many commonly used assays, including FPA, that are limited to artificial substrates.<sup>30</sup> Directed evolution studies would especially benefit from using natural biomass during screening processes. The screening method is of course crucial in this context. As the saying goes, “you get what you screen for”.<sup>4,47</sup> In principle our method of preparing pretreated biomass discs can be used in combination with different sugar readout modes; however, the impact of the substrate on assay results (e.g., unspecific adsorption of dyes) should be carefully assessed.

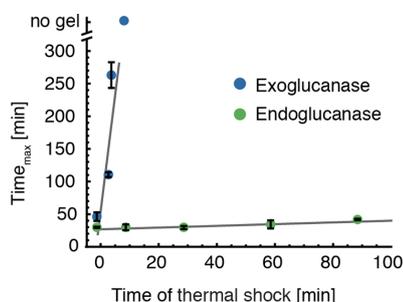
Our label-free HyReS assay is compatible with 96-well plates allowing for easy experiment parallelization and laboratory automation. Liquid handling is relatively uninvolved, and all assay components can be premixed in bulk. After applying HyReS reagents onto biomass discs, no additional pipetting steps are required and readout takes place from the same microtiter plate. This is in contrast to the FPA and other bioenzymatic assays where the addition of further reagents and alteration of buffering conditions is necessary before developing color in an additional incubation step. The general issue of reproducibility and poor comparability due to extreme



**Figure 6.** Detecting synergistic effects between exoglucanases ( $1 \mu\text{M}$ ), endocellulases ( $0.1 \mu\text{M}$ ), and  $\beta$ -glucosidase ( $1 \text{ mg mL}^{-1}$ ). (A and B, top) Changes of epifluorescence signal in time. Shadowed areas represent standard deviation of five measurements. (A and B, bottom) First derivative of fluorescence signal over time. (C) Time at which the peak in fluorescence derivative occurs for various enzyme compositions.

sensitivity to experimental conditions is a widely acknowledged problem for cellulase assays in general.<sup>5,29</sup> Our one-step rapid protocol simplifies the liquid handling and therefore improves reproducibility on any cellulosic substrate of choice. It is also possible to use HyReS system at elevated, more catalytically relevant temperatures (i.e.,  $48 \text{ }^\circ\text{C}$ , data not shown).

Our prior work demonstrated that the same redox/enzyme signaling pathway could be used to polymerize fluorescent hydrogels incorporating a rhodamine-acryl compound.<sup>25</sup> Our



**Figure 7.** Thermostability of cellulases. Time needed to reach maximum of the gel growth rate is plotted against heat-shock time at 80 °C. Linear fits serve as a guide for the eye.

current method significantly differentiates itself from this prior art in several ways. First, the current method is label-free, requiring no dyes whatsoever. Instead we rely on the fluorescent emission inherent to the substrate. Second, we used here a fundamentally different measurement modality based on absorbance/scattering of excitation and emission beams, with a reflective component to the signal contributing in epi-illumination. And third, we have demonstrated for the first time the implementation of a hydrogel-based assay for differentiation of cellulase mixtures based on endo/exo synergy and CBM-targeting ability. Additionally we assayed thermostability of enzymes with the one-pot polymerization assay.

We note the assay as implemented here is primarily a threshold measurement, meaning a certain amount of glucose is required to initiate polymerization. Once the amount of glucose has been produced, polymerization occurs quickly and concludes with relatively little continued gel growth at longer time points. We took as the assay figure of merit the time required to initiate polymerization and found this to be a semiquantitative estimator of hydrolytic enzyme activity.

Despite its advantages, the HyReS system also has some associated limitations. Our one-step protocol introduces possible interference of assay components on cellulolytic activity. In particular, changes in substrate structure and enzyme–substrate interactions induced by PEG40<sup>50</sup> could be of potential concern. However, PEG has been shown to enhance enzymatic hydrolysis of lignocellulose, and we do not expect it to adversely affect most cellulase enzymes.<sup>5,48,49</sup> Potential restrictions on the HyReS assay in terms of pH requirements along with absolute glucose sensitivity are discussed in detail in our previous work.<sup>25</sup>

We also note that due to the complex multistep signal amplification mechanism, the response of our label-free HyReS assay is nonlinear (see Figure 3A). We believe the assay is best suited for determining early stage hydrolytic efficacy, before trapping of enzymes inside the gel structure and transport limitations become dominant. The HyReS assay cannot provide an activity measure in terms of glucose production per unit of time. It is most suitable for applications where direct comparisons between cellulolytic activities at early time points is preferred. However, we do not see this as compromising the assay applicability. Complex synergistic relationships between cellulases and an intricate interplay between substrate structure and enzyme composition limits the predictive power of rational

design for enzymatic cocktails. In most cases a direct comparative empirical approach is indeed necessary.<sup>4</sup>

In conclusion we developed a label-free, polymerization-based HyReS for determining the hydrolysis of lignocellulosic biomass. Through radical polymerization of a cross-linked hydrogel at the location of glucose production, we achieve high signal amplification and specificity for quantifying total cellulolytic activity. Our assay is fast, easy to automate and parallelize, and can be used in combination with arbitrary (ligno)cellulose sources including pretreated biomass. The ability to determine cellulolytic activity, thermostability, exo/endo synergy, and targeting effects in cellulolytic enzyme formulations and cellulosomes establishes the HyReS assay as a valuable method for enzyme screening for improved bio-conversion of lignocellulose.

## EXPERIMENTAL SECTION

**Materials.** Exoglucanase (EXG, cellobiohydrolase I from *Trichoderma longibrachiatum*, specific activity 0.1 U/mg at 40 °C, pH 4.5) and endoglucanase (ENG, endo-1,4- $\beta$ -D-glucanase from *Talaromyces emersonii*, specific activity 64 U/mg at 40 °C, pH 4.5) were purchased from Megazyme (Ireland). Cellulase from *Trichoderma reesei* ATCC 26921 (8 U/mg at 37 °C, pH 5), GOx from *Aspergillus niger*, and  $\beta$ G from almonds (2.1 U/mg at 37 °C, pH 5.0) were purchased from Sigma-Aldrich. Minicellulosomes consisting of three dockerin-containing CelA enzymatic units (cellulase 8A from *Clostridium thermocellum*) arranged on trimodular scaffoldin were purchased from NZYtech (Portugal). Two different scaffoldins, with (3xGH8 + Coh-CBM3-Coh-Coh) and without (3xGH8 + Coh-Coh-Coh) family 3 CBM, were used. Black, flat-bottom polypropylene 96-well plates were purchased from Grenier (Bio-One). All other reagents were obtained from Sigma-Aldrich and used without further purification.

**Biomass Pretreatment.** Two types of energy crops, napier grass (*Pennisetum purpureum*) and miscanthus (*Miscanthus  $\times$  giganteus*), were used as sources of biomass. Plant matter was mechanically processed to produce coarse powder. Non-cellulosic components were extracted with 0.1 M NaOH at 80 °C for 12 h with stirring. After thorough rinsing with water, the biomass sample was delignified in 0.05 M HCl at room temperature for 12 h with stirring. The sample was washed with water until neutral pH was reached. The sample was filtered through Whatman filter paper using Büchner funnel to produce an entangled pad of ~3 mm thickness. The pad was peeled off filter paper and dried overnight at 37 °C. Discs of 6 mm were cut out from the dry, paper-like product using a hole punch.

**Raman Spectroscopy.** Raman spectra were obtained using T64000 triple grating Raman system (Horiba Scientific, France). The measurements were performed in air using a 568 nm argon/krypton gas laser line (Coherent) and a 100 $\times$  MPlanN air objective (NA 0.9, Olympus). Spectra were calibrated with the Raman line of silicon at 520.70  $\text{cm}^{-1}$ .

**HyReS Assay.** All measurements were performed in 20 mM sodium acetate (NaAc) buffer at pH 4.5. The HyReS mix supplemented with cellulolytic enzymes of interest was freshly prepared before each experiment and preheated to 37 °C. Composition of the standard reagent mixture is shown in Table 1.

A black 96-well polypropylene plate with flat bottom was first cleaned with isopropyl alcohol and washed with deionized water. The biomass discs were carefully placed at the bottom of the plate wells, and the plate was preheated to 37 °C. The wells

Table 1. Components of Label-Free HyReS Assay

component	concentration
glucose oxidase	1 mg mL <sup>-1</sup>
FeSO <sub>4</sub>	250 μM
ascorbic acid	250 μM
PEG diacrylate (M <sub>n</sub> 575)	150 mg mL <sup>-1</sup>
NaAc buffer, pH 4.5	20 mM

were then filled with 200 μL of HyReS components and cellulase mixture using a multipipette and the plate was put into a multiwell plate reader (Infinite M1000 Pro, Tecan). During incubation at 37 °C the fluorescence intensity was measured from the top using a time-resolved kinetic cycle. The excitation wavelength of 365 nm and emission wavelength of 430 nm were used, and 16 reads on 4 × 4 grid were performed in each well.

**Data Analysis.** Each experiment was performed in quintuplicate, and a mean autofluorescence  $f(t)$  with standard deviation  $\sigma_f(t)$  was determined. Normalized autofluorescence  $F(t)$  was calculated with respect to fluorescence at the beginning of the experiment  $F(t) = f(t)/f(0)$ . Error bars are plotted as standard deviation of the normalized autofluorescence  $\sigma_F(t)$ . Prior to numerical differentiation data was smoothed using moving average function in Igor Pro software package (Wavemetrics) using box sizes  $(2M + 1)$  between 20 and 200. It is important to notice that smoothed curves were only used for numerical differentiation of data. Plots showing changes of fluorescence in time in the manuscript represent original, nonsmoothed data.

The time at which a maximum in the differentiated data occurred  $t_{\max}$  was used for assessing cellulolytic activity of assay enzymes. It is reported with an error  $\sigma_{t_{\max}}$  calculated from  $\sigma_F(t_{\max})$  according to the following formula:

$$\sigma_{t_{\max}} = \left( \left. \frac{dF(t)}{dt} \right|_{t=t_{\max}} \right)^{-1} \sigma_F(t_{\max})$$

**Thermostability Measurement.** A 10 μM solution of EXG/ENG in NaAc was heat-shocked at 80 °C for up to 90 min. Afterward it was cooled to room temperature and mixed with HyReS reagents to obtain detection solutions containing 2 μM EDG. Cellulolytic activity assay was performed as described above.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: michael.nash@lmu.de.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Miscanthus samples were donated by Professor Pude (Field Lab Campus Klein-Altendorf, University of Bonn). We thank Patrick Urban (Professor Feldmann, LMU Munich) for help with recording Raman spectra. The authors acknowledge funding from the EU Seventh Framework Program NMP4-SL-2013-604530 (CellulosomePlus) and the Nanosystems Initiative Munich. M.A.N. acknowledges funding from Society in Science—The Branco Weiss Fellowship administered by the Swiss Federal Institute of Technology (ETH Zurich).

## REFERENCES

- (1) Heinzelman, P.; Snow, C. D.; Wu, L.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5610–5615.
- (2) Decker, S. R.; Brunecky, R.; Tucker, M. P.; Himmel, M. E.; Selig, M. J. *Bioenergy Res.* **2009**, *2*, 179–192.
- (3) Song, L.; Siguier, B.; Dumon, C.; Bozonnet, S.; O'Donohue, M. J. *Biotechnol. Biofuels* **2011**, *5*, 3–3.
- (4) Zhang, Y. P.; Himmel, M. E.; Mielenz, J. R. *Biotechnol. Adv.* **2006**, *24*, 452–481.
- (5) Bailey, M. J.; Biely, P.; Poutanen, K. J. *Biotechnol.* **1991**, *23*, 257–270.
- (6) Luterbacher, J. S.; Moran-Mirabal, J. M.; Burkholder, E. W.; Walker, L. P. *Biotechnol. Bioeng.* **2015**, *112*, 32–42.
- (7) Yang, D.; Moran-Mirabal, J. M.; Parlange, J.-Y.; Walker, L. P. *Biotechnol. Bioeng.* **2013**, *110*, 2836–2845.
- (8) Paës, G. *Molecules* **2014**, *19*, 9380–9402.
- (9) Ding, S.-Y.; Liu, Y. S.; Zeng, Y.; Himmel, M. E.; Baker, J. O.; Bayer, E. A. *Science* **2012**, *338*, 1055–1060.
- (10) Eibinger, M.; Bubner, P.; Ganner, T.; Plank, H.; Nidetzky, B. *FEBS J.* **2013**, *281*, 275–290.
- (11) Bubner, P.; Plank, H.; Nidetzky, B. *Biotechnol. Bioeng.* **2013**, *110*, 1529–1549.
- (12) Santa-Maria, M.; Jeoh, T. *Biomacromolecules* **2010**, *11*, 2000–2007.
- (13) Somerville, C.; Youngs, H.; Taylor, C.; Davis, S. C.; Long, S. P. *Science* **2010**, *329*, 790–792.
- (14) Miller, G. L. *Anal. Chem.* **1959**, *31*, 426–428.
- (15) Ghose, T. K. *Pure Appl. Chem.* **1987**, *59*, 257–268.
- (16) Decker, S. R.; Adney, W. S.; Jennings, E.; Vinzant, T. B.; Himmel, M. E. *Appl. Biochem. Biotechnol.* **2003**, *105–08*, 689–703.
- (17) Xiao, Z.; Storms, R.; Tsang, A. *Biotechnol. Bioeng.* **2004**, *88*, 832–837.
- (18) Cianchetta, S.; Galletti, S.; Burzi, P. L.; Cerato, C. *Biotechnol. Bioeng.* **2010**, *107*, 461–468.
- (19) Kongruang, S.; Bothwell, M. K.; McGuire, J.; Zhou, M.; Haugland, R. P. *Enzyme Microb. Technol.* **2003**, *32*, 539–545.
- (20) Chandrasekaran, A.; Bharadwaj, R.; Park, J. I.; Sapra, R.; Adams, P. D.; Singh, A. K. *J. Proteome Res.* **2010**, *9*, 5677–5683.
- (21) Wright, W. R.; Rainwater, J. C.; Tolle, L. D. *Clin. Chem.* **1971**, *17*, 1010–1015.
- (22) Chundawat, S. P. S.; Balan, V.; Dale, B. E. *Biotechnol. Bioeng.* **2008**, *99*, 1281–1294.
- (23) Chang, C.; Sutarich, J.; Bharadwaj, R.; Chandrasekaran, A.; Adams, P. D.; Singh, A. K. *Lab Chip* **2013**, *13*, 1817–1822.
- (24) Ostafe, R.; Prodanovic, R.; Commandeur, U.; Fischer, R. *Anal. Biochem.* **2013**, *435*, 93–98.
- (25) Malinowska, K.; Verdorfer, T.; Meinhold, A.; Milles, L. F.; Funk, V.; Gaub, H. E.; Nash, M. A. *ChemSusChem* **2014**, *7*, 2825–2831.
- (26) Plitt, K. F.; Toner, S. D. *J. Appl. Polym. Sci.* **1961**, *5*, 534–538.
- (27) Olmstead, J. A.; Gray, D. G. *J. Photochem. Photobiol., A* **1993**, *73*, 59–65.
- (28) Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68–75.
- (29) Dashtban, M.; Maki, M.; Leung, K. T.; Mao, C.; Qin, W. *Crit. Rev. Biotechnol.* **2010**, *30*, 302–309.
- (30) Kabel, M. A.; van der Maarel, M. J. E. C.; Klip, G.; Voragen, A. G. J.; Schols, H. A. *Biotechnol. Bioeng.* **2005**, *93*, 56–63.
- (31) Dutta, S.; Wu, K. C. W. *Green Chem.* **2014**, *16*, 4615–4626.
- (32) Bornscheuer, U.; Buchholz, K.; Seibel, J. *Angew. Chem., Int. Ed.* **2014**, *53*, 10876–10893.
- (33) King, B. C.; Donnelly, M. K.; Bergstrom, G. C.; Walker, L. P.; Gibson, D. M. *Biotechnol. Bioeng.* **2009**, *102*, 1033–1044.
- (34) Alvira, P.; Negro, M. J.; Sáez, F.; Ballesteros, M. J. *Chem. Technol. Biotechnol.* **2010**, *85*, 1291–1297.
- (35) Johnson, L. M.; DeForest, C. A.; Pendurti, A.; Anseth, K. S.; Bowman, C. N. *ACS Appl. Mater. Interfaces* **2010**, *2*, 1963–1972.
- (36) Pitzler, C.; Wirtz, G.; Vojcic, L.; Hiltl, S.; Böker, A.; Martinez, R.; Schwaneberg, U. *Chem. Biol.* **2014**, *21*, 1733–1742.

- (37) Xu, J.; Jordan, R. B. *Inorg. Chem.* **1990**, *29*, 4180–4184.
- (38) Agarwal, U. P.; Sally, R. A. *Appl. Spectrosc.* **1997**, *51*, 1648–1655.
- (39) Zollinger, H. *Color Chemistry: Syntheses, Properties, and Applications of Organic Dyes and Pigments*; Wiley-VCH: Weinheim, Germany, 2003.
- (40) Kostylev, M.; Wilson, D. *Biofuels* **2012**, *3*, 61–70.
- (41) Murashima, K.; Kosugi, A.; Doi, R. H. *J. Bacteriol.* **2002**, *184*, 5088–5095.
- (42) Andersen, N.; Johansen, K. S.; Michelsen, M.; Stenby, E. H.; Krogh, K. B. R. M.; Olsson, L. *Enzyme Microb. Technol.* **2008**, *42*, 362–370.
- (43) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. *Biochem. J.* **2004**, *382*, 769–781.
- (44) Carrard, G.; Koivula, A.; Söderlund, H.; Béguin, P. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10342–10347.
- (45) Fierobe, H.-P.; Bayer, E. A.; Tardif, C.; Czjzek, M.; Mechaly, A.; Bélaïch, A.; Lamed, R.; Shoham, Y.; Bélaïch, J.-P. *J. Biol. Chem.* **2002**, *277*, 49621–49630.
- (46) Blumer-Schuette, S. E.; Brown, S. D.; Sander, K. B.; Bayer, E. A.; Kataeva, I.; Zurawski, J. V.; Conway, J. M.; Adams, M. W.; Kelly, R. M. *FEMS Microbiol. Rev.* **2014**, *38*, 393–448.
- (47) Schmidt-Dannert, C.; Arnold, F. H. *Trends Biotechnol.* **1999**, *17*, 135–136.
- (48) Helle, S. S.; Duff, S. J. B.; Cooper, D. G. *Biotechnol. Bioeng.* **1993**, *42*, 611–617.
- (49) Kumar, R.; Wyman, C. E. *Biotechnol. Bioeng.* **2009**, *102*, 1544–1557.
- (50) Li, J.; Li, S.; Fan, C.; Yan, Z. *Colloids Surf. B* **2012**, *89*, 203–210.





**Part IV**

**Conclusion**



---

Nature is rich in a wide variety of complex, synergistic, and highly functional protein-based multicomponent assemblies. This thesis contributes to a more complete understanding of one of them - the cellulosome - and provides new strategies to improve rationally designed artificial enzyme networks, such as designer cellulosomes.

Not only do cellulosomes hydrolyze lignocellulosic biomass with high efficiency, but furthermore anchor their host cells to their nutrition source *via* CBM-bearing scaffoldins. Besides the CBMs, scaffoldins contain several different cohesin domains with high sequence similarity. I was able to show that the mechanical stability of cohesin domains strongly correlates with their position within the scaffoldin. The so-called bridging cohesins, which transduce force between the CBM and the cell's surface, are able to withstand notably higher unfolding forces than their loose hanging counterparts. This fact may be explained by an evolutionary pressure, since mechanically weak bridging cohesins would be unable to remain folded and thus bind enzyme-bearing dockerins under mechanical stress, resulting in inefficient cellulose degradation and a disadvantage for the host.

The intriguingly strong stabilizing effect of dockerin binding on cohesin mechanostability further contributes to the understanding of cellulosome architecture and sparks new questions about its mechanobiology. For example, through which mechanism does dockerin binding influence the cohesins mechanostability, given that the cohesin's binding site is located on the opposite side of its region responsible for mechanostability? Do different dockerins have differing stabilizing effects? And does the stabilizing effect correlate with binding affinity, energy or strength?

I was also able to directly apply results gained from steered molecular dynamics simulations to enhance the mechanical robustness of the weakest tested cohesin domain. The ability to correctly predict the effect of single amino acid mutations on fold strength and to tune the robustness of proteins reveals the potential of the applied *in silico* protocols to design future mutants.

The results in this thesis contribute to the understanding of cellulosome mechanobiology and that the presented methods open the door to the design of enzyme networks with high mechanical rigidity, an aspect so far not taken into account in the development of designer cellulosomes. The ability to choose from cohesin-dockerin pairs with high affinity, high unbinding force and high mechanical stability makes the modular cellulosomal system very attractive for all kinds of molecular assembly applications across many biotechnological fields, both *in vivo* and *in vitro*.

The polymerization-based cellulase assay for the quantification and localization of enzymatic cellulose degradation is furthermore promising for high-throughput screenings of multi-component enzyme mixtures and designer cellulosomes across various biomass substrates. Moreover, the possibility to image cellulose hydrolysis in real time and to pinpoint enzymatic activity to topological cellulose features might reveal bottlenecks in biomass breakdown.

The insights into the mechanobiology of the cellulosome and the tool-set to investigate the turnover of varying enzyme combinations will hopefully contribute to the development of a new generation of designer cellulosomes, ore more generally: rationally designed enzyme networks.



**Part V**

**Appendix**



# List of Publications

- [1] **T. Verdorfer** and H. E. Gaub  
“Ligand Binding Stabilizes Cellulosomal Cohesins as Revealed by AFM-based Single-Molecule Force Spectroscopy”  
Scientific Reports, 2018  
DOI: 10.1038/s41598-018-27085-x
- [2] **T. Verdorfer**, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash and H. E. Gaub  
“Combining *in Vitro* and *in Silico* Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics”  
Journal of the American Chemical Society, 2017  
DOI: 10.1021/jacs.7b07574
- [3] C. Schoeler, **T. Verdorfer**, H. E. Gaub and M. A. Nash  
“Biasing Effects of Receptor-Ligand Complexes on Protein-Unfolding Statistics”  
Physical Review E, 2016  
DOI: 10.1103/PhysRevE.94.042412
- [4] K. H. Malinowska, T. Rind, **T. Verdorfer**, H. E. Gaub and M. A. Nash  
“Quantifying Synergy, Thermostability, and Targeting of Cellulolytic Enzymes and Cellulosomes with Polymerization-Based Amplification”  
Analytical Chemistry, 2015  
DOI: 10.1021/acs.analchem.5b00936
- [5] M. Otten\*, W. Ott\*, M. A. Jobst\*, L. F. Milles, **T. Verdorfer**, D. A. Pippig, M. A. Nash, and H. E. Gaub  
“From Genes to Protein Mechanics on a Chip”  
Nature Methods, 2014  
DOI: 10.1038/nmeth.3099
- [6] K. H. Malinowska\*, **T. Verdorfer**\*, A. Meinhold, L. F. Milles, V. Funk, H. E. Gaub and M. A. Nash  
“Redox-Initiated Hydrogel System for Detection and Real-Time Imaging of Cellulolytic Enzyme Activity”  
ChemSusChem, 2014  
DOI: 10.1002/cssc.201402796

---

\* These authors contributed equally to this work



# List of Figures

1.1	Structure of lignocellulose . . . . .	10
1.2	The two paradigms of cellulase organization . . . . .	11
1.3	Simplified architectural model of <i>A. cellulolyticus</i> ' cellulosome system	13
2.1	Schematic of an AFM in single-molecule force spectroscopy mode .	20



# Bibliography

- [1] M. Meselson and F. W. Stahl, "The replication of dna in escherichia coli", *Proceedings of the National Academy of Sciences*, vol. 44, no. 7, pp. 671–682, 1958 (cit. on p. 3).
- [2] B. B. Kahn, T. Alquier, D. Carling, and D. G. Hardie, "AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism", *Cell Metabolism*, vol. 1, no. 1, pp. 15–25, Jan. 2005. doi: 10.1016/j.cmet.2004.12.003 (cit. on p. 3).
- [3] K. Simons and D. Toomre, "Lipid rafts and signal transduction", *Nature Reviews Molecular Cell Biology*, vol. 1, no. 1, pp. 31–39, Sep. 2000. doi: 10.1038/35036052 (cit. on p. 3).
- [4] R. Mahley, "Apolipoprotein e: Cholesterol transport protein with expanding role in cell biology", *Science*, vol. 240, no. 4852, pp. 622–630, Apr. 1988. doi: 10.1126/science.3283935 (cit. on p. 3).
- [5] J. Howard *et al.*, "Mechanics of motor proteins and the cytoskeleton", 2001 (cit. on p. 3).
- [6] E. A. Bayer, J.-P. Belaich, Y. Shoham, and R. Lamed, "The cellulosomes: Multienzyme machines for degradation of plant cell wall polysaccharides", *Annu. Rev. Microbiol.*, vol. 58, pp. 521–554, 2004. doi: 10.1146/annurev.micro.57.030502.091022 (cit. on p. 3).
- [7] A. L. Carvalho, F. M. V. Dias, J. A. M. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. A. Ferreira, M. J. Romao, and C. M. G. A. Fontes, "Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex", *Proceedings of the National Academy of Sciences*, vol. 100, no. 24, pp. 13 809–13 814, Nov. 2003. doi: 10.1073/pnas.1936124100 (cit. on p. 3).
- [8] S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, and H. E. Gaub, "Single-molecule dissection of the high-affinity cohesin-dockerin complex", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 50, pp. 20 431–20 436, 2012. doi: 10.1073/pnas.1211929109 (cit. on pp. 3, 11, 15).
- [9] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, and M. A. Nash, "Ultrastable cellulosome-adhesion complex tightens under load", *Nat. Commun.*, vol. 5, no. 5635, pp. 1–8, 2014. doi: 10.1038/ncomms6635 (cit. on pp. 3, 11, 15).
- [10] M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub, and M. A. Nash, "Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy", *Elife*, vol. 4, e10319, 2015. doi: 10.7554/eLife.10319 (cit. on pp. 3, 15).

- [11] A. Valbuena, J. Oroz, R. Hervás, A. M. Vera, D. Rodríguez, M. Menéndez, J. I. Sulkowska, M. Cieplak, and M. Carrión-Vázquez, "On the remarkable mechanostability of scaffoldins and the mechanical clamp motif", *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 33, pp. 13 791–13 796, 2009. doi: 10.1073/pnas.0813093106 (cit. on pp. 3, 14, 29).
- [12] H. Chen, *Biotechnology of Lignocellulose*. Springer Netherlands, 2014. doi: 10.1007/978-94-007-6898-7 (cit. on p. 9).
- [13] C. Somerville, "Toward a systems approach to understanding plant cell walls", *Science*, vol. 306, no. 5705, pp. 2206–2211, Dec. 2004. doi: 10.1126/science.1102765 (cit. on p. 9).
- [14] U. Bornscheuer, K. Buchholz, and J. Seibel, "Enzymatic degradation of (ligno)cellulose", *Angewandte Chemie International Edition*, vol. 53, no. 41, pp. 10 876–10 893, Aug. 2014. doi: 10.1002/anie.201309953 (cit. on p. 9).
- [15] J. Wang, J. Xi, and Y. Wang, "Recent advances in the catalytic production of glucose from lignocellulosic biomass", *Green Chemistry*, vol. 17, no. 2, pp. 737–751, 2015. doi: 10.1039/c4gc02034k (cit. on p. 9).
- [16] B. Mahro and M. Timm, "Potential of biowaste from the food industry as a biomass resource", *Engineering in Life Sciences*, vol. 7, no. 5, pp. 457–468, Oct. 2007. doi: 10.1002/elsc.200620206 (cit. on p. 10).
- [17] N. MOSIER, "Features of promising technologies for pretreatment of lignocellulosic biomass", *Bioresource Technology*, vol. 96, no. 6, pp. 673–686, Apr. 2007. doi: 10.1016/j.biortech.2004.06.025 (cit. on p. 10).
- [18] B. Seiboth, C. Ivanova, and V. Seidl-Seiboth, "Trichoderma reesei: A fungal enzyme producer for cellulosic biofuels", in *Biofuel Production-Recent Developments and Prospects*, InTech, Sep. 2011. doi: 10.5772/16848 (cit. on p. 10).
- [19] R. Lamed, E. Setter, and E. A. Bayer, "Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum", *J. Bacteriol.*, vol. 156, no. 2, pp. 828–836, 1983 (cit. on p. 10).
- [20] Y. Hamberg, V. Ruimy-Israeli, B. Dassa, Y. Barak, R. Lamed, K. Cameron, C. M. Fontes, E. A. Bayer, and D. B. Fried, "Elaborate cellulosome architecture of acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions", *PeerJ*, vol. 2, e636, Sep. 2014. doi: 10.7717/peerj.636 (cit. on pp. 10, 13).
- [21] S. Morais, Y. Barak, R. Lamed, D. B. Wilson, Q. Xu, M. E. Himmel, and E. A. Bayer, "Paradigmatic status of an endo- and exoglucanase and its effect on crystalline cellulose degradation", *Biotechnology for Biofuels*, vol. 5, no. 1, p. 78, 2012. doi: 10.1186/1754-6834-5-78 (cit. on pp. 11, 16).
- [22] D. Guillen, S. Sanchez, and R. Rodriguez-Sanoja, "Carbohydrate-binding domains: Multiplicity of biological roles", *Applied Microbiology and Biotechnology*, vol. 85, no. 5, pp. 1241–1249, Nov. 2009. doi: 10.1007/s00253-009-2331-y (cit. on p. 11).

- 
- [23] R. Haimovitz, Y. Barak, E. Morag, M. Voronov-Goldman, Y. Shoham, R. Lamed, and E. A. Bayer, "Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules", *PROTEOMICS*, vol. 8, no. 5, pp. 968–979, Mar. 2008. doi: 10.1002/pmic.200700486 (cit. on p. 11).
- [24] I. Noach, R. Lamed, Q. Xu, S. Rosenheck, L. J. W. Shimon, E. A. Bayer, and F. Frolow, "Preliminary x-ray characterization and phasing of a type II cohesin domain from the cellulosome of *Acetivibrio cellulolyticus*", *Acta Crystallographica Section D Biological Crystallography*, vol. 59, no. 9, pp. 1670–1673, Aug. 2003. doi: 10.1107/s0907444903014094 (cit. on p. 11).
- [25] A. Mechaly, H.-P. Fierobe, A. Belaich, J.-P. Belaich, R. Lamed, Y. Shoham, and E. A. Bayer, "Cohesin-dockerin interaction in cellulosome assembly", *Journal of Biological Chemistry*, vol. 276, no. 13, pp. 9883–9888, Jan. 2001. doi: 10.1074/jbc.M009237200 (cit. on p. 11).
- [26] A. L. Carvalho, F. M. V. Dias, T. Nagy, J. A. M. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. A. Ferreira, M. J. Romao, C. M. G. A. Fontes, and H. J. Gilbert, "Evidence for a dual binding mode of dockerin modules to cohesins", *Proceedings of the National Academy of Sciences*, vol. 104, no. 9, pp. 3089–3094, Feb. 2007. doi: 10.1073/pnas.0611173104 (cit. on p. 11).
- [27] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, and E. A. Bayer, "Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses", *J. Biol. Chem.*, vol. 288, no. 23, pp. 16 827–16 838, 2013. doi: 10.1074/jbc.M113.466672 (cit. on p. 11).
- [28] E. T. Reese, "Enzymatic hydrolysis of cellulose", *Applied Microbiology*, vol. 4, no. 1, pp. 39–45, 1956 (cit. on p. 12).
- [29] F. Allen, R. Andreotti, D. E. Eveleigh, and J. Nystrom, "Mary elizabeth hickox mandels, 90, bioenergy leader", *Biotechnology for Biofuels*, vol. 2, no. 1, p. 22, 2009. doi: 10.1186/1754-6834-2-22 (cit. on p. 12).
- [30] F.-W. Bai, C.-G. Liu, H. Huang, and G. T. Tsao, Eds., *Biotechnology in China III: Biofuels and Bioenergy*. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-28478-6 (cit. on p. 12).
- [31] C. M. Payne, B. C. Knott, H. B. Mayes, H. Hansson, M. E. Himmel, M. Sandgren, J. Stahlberg, and G. T. Beckham, "Fungal cellulases", *Chemical Reviews*, vol. 115, no. 3, pp. 1308–1448, Jan. 2015. doi: 10.1021/cr500351c (cit. on p. 12).
- [32] M. Paloheimo, T. Haarmann, S. Mäkinen, and J. Vehmaanperä, "Production of industrial enzymes in *trichoderma reesei*", in *Fungal Biology*, Springer International Publishing, 2016, pp. 23–57. doi: 10.1007/978-3-319-27951-0\_2 (cit. on p. 12).
-

- [33] D. Martinez, R. M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. E. Baker, J. Chapman, O. Chertkov, P. M. Coutinho, D. Cullen, E. G. J. Danchin, I. V. Grigoriev, P. Harris, M. Jackson, C. P. Kubicek, C. S. Han, I. Ho, L. F. Larrondo, A. L. de Leon, J. K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A. A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. L. Schoch, J. Yao, R. Barabote, M. A. Nelson, C. Detter, D. Bruce, C. R. Kuske, G. Xie, P. Richardson, D. S. Rokhsar, S. M. Lucas, E. M. Rubin, N. Dunn-Coleman, M. Ward, and T. S. Brettin, "Genome sequencing and analysis of the biomass-degrading fungus *trichoderma reesei* (syn. *hypocrea jecorina*)", *Nature Biotechnology*, vol. 26, no. 5, pp. 553–560, May 2008. doi: 10.1038/nbt1403 (cit. on p. 12).
- [34] S. Shoemaker, V. Schweickart, M. Ladner, D. Gelfand, S. Kwok, K. Myambo, and M. Innis, "Molecular cloning of exo-cellobiohydrolase i derived from *trichoderma reesei* strain 127", *Nature Biotechnology*, vol. 1, no. 8, pp. 691–696, Oct. 1983. doi: 10.1038/nbt1083-691 (cit. on p. 12).
- [35] C. Divne, J. Stahlberg, T. Reinikainen, L. Ruohonen, G. Pettersson, J. Knowles, T. Teeri, and T. Jones, "The three-dimensional crystal structure of the catalytic core of cellobiohydrolase i from *trichoderma reesei*", *Science*, vol. 265, no. 5171, pp. 524–528, Jul. 1994. doi: 10.1126/science.8036495 (cit. on p. 12).
- [36] K. Igarashi, T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttilä, T. Ando, and M. Samejima, "Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface", *Science*, vol. 333, no. 6047, pp. 1279–1282, Sep. 2011. doi: 10.1126/science.1208386 (cit. on p. 12).
- [37] H. Durand, M. Clanet, and G. Tiraby, "Genetic improvement of *trichoderma reesei* for large scale cellulase production", *Enzyme and Microbial Technology*, vol. 10, no. 6, pp. 341–346, Jun. 1988. doi: 10.1016/0141-0229(88)90012-9 (cit. on p. 12).
- [38] R. Peterson and H. Nevalainen, "Trichoderma reesei RUT-c30 - thirty years of strain improvement", *Microbiology*, vol. 158, no. 1, pp. 58–68, Oct. 2011. doi: 10.1099/mic.0.054031-0 (cit. on p. 12).
- [39] R. F. H. Dekker, "Bioconversion of hemicellulose: Aspects of hemicellulase production by *Trichoderma reesei* QM 9414 and enzymic saccharification of hemicellulose", *Biotechnology and Bioengineering*, vol. 25, no. 4, pp. 1127–1146, Apr. 1983. doi: 10.1002/bit.260250419 (cit. on p. 12).
- [40] P. Biely, J. Puls, and H. Schneider, "Acetyl xylan esterases in fungal cellulolytic systems", *FEBS Letters*, vol. 186, no. 1, pp. 80–84, Jul. 1985. doi: 10.1016/0014-5793(85)81343-0 (cit. on p. 12).
- [41] M. Saloheimo, P. Lehtovaara, M. Penttilä, T. Teeri, J. Ståhlberg, G. Johansson, G. Pettersson, M. Claeysens, P. Tomme, and J. Knowles, "EGIII, a new endoglucanase from *trichoderma reesei*: The characterization of both gene and enzyme", *Gene*, vol. 63, no. 1, pp. 11–21, Mar. 1988. doi: 10.1016/0378-1119(88)90541-0 (cit. on p. 12).

- [42] A. Törrönen, R. L. Mach, R. Messner, R. Gonzalez, N. Kalkkinen, A. Harkki, and C. P. Kubicek, "The two major xylanases from *Trichoderma reesei*: Characterization of both enzymes and genes", *Nature Biotechnology*, vol. 10, no. 11, pp. 1461–1465, Nov. 1992. doi: 10.1038/nbt1192-1461 (cit. on p. 12).
- [43] M. Tenkanen, J. Buchert, and L. Viikari, "Binding of hemicellulases on isolated polysaccharide substrates", *Enzyme and Microbial Technology*, vol. 17, no. 6, pp. 499–505, Jun. 1995. doi: 10.1016/0141-0229(94)00050-2 (cit. on p. 12).
- [44] M. Tenkanen, E. Luonteri, and A. Teleman, "Effect of side groups on the action of beta-xylosidase from *Trichoderma reesei* against substituted xylo-oligosaccharides", *FEBS Letters*, vol. 399, no. 3, pp. 303–306, Dec. 1996. doi: 10.1016/S0014-5793(96)01313-0 (cit. on p. 12).
- [45] P. Biely, G. Côté, L. Kremnický, R. Greene, and M. Tenkanen, "Action of acetylxylan esterase from *Trichoderma reesei* on acetylated methyl glycosides", *FEBS Letters*, vol. 420, no. 2-3, pp. 121–124, Dec. 1997. doi: 10.1016/S0014-5793(97)01500-7 (cit. on p. 12).
- [46] M. Tenkanen, M. Vršanská, M. Siika-aho, D. W. Wong, V. Puchart, M. Penttilä, M. Saloheimo, and P. Biely, "Xylanase XYN IV from *Trichoderma reesei* showing exo- and endo-xylanase activity", *FEBS Journal*, vol. 280, no. 1, pp. 285–301, Dec. 2012. doi: 10.1111/febs.12069 (cit. on p. 12).
- [47] P. Biely, M. Cizšárová, J. W. Agger, X.-L. Li, V. Puchart, M. Vršanská, V. G. Eijsink, and B. Westereng, "Trichoderma reesei CE16 acetyl esterase and its role in enzymatic degradation of acetylated hemicellulose", *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1840, no. 1, pp. 516–525, Jan. 2014. doi: 10.1016/j.bbagen.2013.10.008 (cit. on p. 12).
- [48] P. Biely, V. Puchart, M. A. Stringer, and K. B. M. Krogh, "Trichoderma reesei XYN VI - a novel appendage-dependent eukaryotic glucuronoxylan hydrolase", *FEBS Journal*, vol. 281, no. 17, pp. 3894–3903, Jul. 2014. doi: 10.1111/febs.12925 (cit. on p. 12).
- [49] M. Mandels and E. T. Reese, "Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals", *Journal of Bacteriology*, vol. 73, no. 2, pp. 269–278, 1957 (cit. on p. 12).
- [50] M. Mandels and E. T. Reese, "Induction of cellulase in fungi by cellobiose", *Journal of Bacteriology*, vol. 79, no. 6, pp. 816–826, 1960 (cit. on p. 12).
- [51] M. Mandels, F. W. Parrish, and E. T. Reese, "Sophorose as an inducer of cellulase in *Trichoderma viride*", *Journal of Bacteriology*, vol. 83, pp. 400–408, 1962 (cit. on p. 12).
- [52] R. H. Bischof, J. Ramoni, and B. Seiboth, "Cellulases and beyond: The first 70 years of the enzyme producer *Trichoderma reesei*", *Microbial Cell Factories*, vol. 15, no. 1, Jun. 2016. doi: 10.1186/s12934-016-0507-6 (cit. on p. 12).
- [53] A. W. KHAN, "Cellulolytic enzyme system of *Acetivibrio cellulolyticus*, a newly isolated anaerobe", *Microbiology*, vol. 121, no. 2, pp. 499–502, Dec. 1980. doi: 10.1099/00221287-121-2-499 (cit. on p. 13).

- [54] P. R. Pandya, K. M. Singh, S. Parnerkar, A. K. Tripathi, H. H. Mehta, D. N. Rank, R. K. Kothari, and C. G. Joshi, "Bacterial diversity in the rumen of indian surti buffalo (*bubalus bubalis*), assessed by 16s rDNA analysis", *Journal of Applied Genetics*, vol. 51, no. 3, pp. 395–402, Sep. 2010. doi: 10.1007/bf03208869 (cit. on p. 13).
- [55] G. B. PATEL, A. W. KHAN, B. J. AGNEW, and J. R. COLVIN, "Isolation and characterization of an anaerobic, cellulolytic microorganism, *acetivibrio cellulolyticus* gen. nov., sp. nov. ", *International Journal of Systematic Bacteriology*, vol. 30, no. 1, pp. 179–185, Jan. 1980. doi: 10.1099/00207713-30-1-179 (cit. on p. 13).
- [56] C. Mackenzie, D. Bilous, and G. Patel, "Studies on cellulose hydrolysis by *acetivibrio cellulolyticus*", *Applied and environmental microbiology*, vol. 50, pp. 243–8, Sep. 1985 (cit. on p. 13).
- [57] Q. Xu, Y. Barak, R. Kenig, Y. Shoham, E. A. Bayer, and R. Lamed, "A novel *acetivibrio cellulolyticus* anchoring scaffoldin that bears divergent cohesins", *Journal of Bacteriology*, vol. 186, no. 17, pp. 5782–5789, Aug. 2004. doi: 10.1128/jb.186.17.5782-5789.2004 (cit. on p. 13).
- [58] I. Noach, M. Levy-Assaraf, R. Lamed, L. J. Shimon, F. Frolow, and E. A. Bayer, "Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad", *Journal of Molecular Biology*, vol. 399, no. 2, pp. 294–305, Jun. 2010. doi: 10.1016/j.jmb.2010.04.013 (cit. on p. 13).
- [59] B. Dassa, I. Borovok, R. Lamed, B. Henrissat, P. Coutinho, C. L. Hemme, Y. Huang, J. Zhou, and E. A. Bayer, "Genome-wide analysis of *acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system", *BMC Genomics*, vol. 13, no. 1, p. 210, 2012. doi: 10.1186/1471-2164-13-210 (cit. on p. 13).
- [60] C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash, and H. E. Gaub, "Mapping mechanical force propagation through biomolecular complexes", *Nano Lett.*, vol. 15, no. 11, pp. 7370–7376, 2015. doi: 10.1021/acs.nanolett.5b02727 (cit. on p. 15).
- [61] E. Durner, W. Ott, M. A. Nash, and H. E. Gaub, "Post-Translational Sortase-Mediated attachment of High-Strength force spectroscopy handles", *ACS Omega*, vol. 2, no. 6, pp. 3064–3069, 2017. doi: 10.1021/acsomega.7b00478 (cit. on p. 15).
- [62] W. Ott, M. A. Jobst, M. S. Bauer, E. Durner, L. F. Milles, M. A. Nash, and H. E. Gaub, "Elastin-like polypeptide linkers for Single-Molecule force spectroscopy", *ACS Nano*, vol. 11, no. 6, pp. 6346–6354, 2017. doi: 10.1021/acsnano.7b02694 (cit. on p. 15).
- [63] F. Sumbul, A. Marchesi, and F. Rico, "History, rare, and multiple events of mechanical unfolding of repeat proteins", *The Journal of Chemical Physics*, vol. 148, no. 12, p. 123335, Mar. 2018. doi: 10.1063/1.5013259 (cit. on p. 15).

- 
- [64] E. A. Bayer, E. Morag, and R. Lamed, "The cellulosome - a treasure-trove for biotechnology", *Trends in Biotechnology*, vol. 12, no. 9, pp. 379–386, Sep. 1994. doi: 10.1016/0167-7799(94)90039-6 (cit. on p. 15).
- [65] H.-P. Fierobe, E. A. Bayer, C. Tardif, M. Czjzek, A. Mechaly, A. Belaich, R. Lamed, Y. Shoham, and J.-P. Belaich, "Degradation of cellulose substrates by cellulosome chimeras", *Journal of Biological Chemistry*, vol. 277, no. 51, pp. 49 621–49 630, Oct. 2002. doi: 10.1074/jbc.m207672200 (cit. on p. 15).
- [66] S.-L. Tsai, N. A. DaSilva, and W. Chen, "Functional display of complex cellulosomes on the yeast surface via adaptive assembly", *ACS Synthetic Biology*, vol. 2, no. 1, pp. 14–21, Jul. 2012. doi: 10.1021/sb300047u (cit. on p. 16).
- [67] G. Gefen, M. Anbar, E. Morag, R. Lamed, and E. A. Bayer, "Enhanced cellulose degradation by targeted integration of a cohesin-fused  $\beta$ -glucosidase into the clostridium thermocellum cellulosome", *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. 10 298–10 303, Jun. 2012. doi: 10.1073/pnas.1202747109 (cit. on p. 16).
- [68] H.-Y. Cho, H. Yukawa, M. Inui, R. H. Doi, and S.-L. Wong, "Production of minicellulosomes from clostridium cellulovorans in bacillus subtilis WB800", *Applied and Environmental Microbiology*, vol. 70, no. 9, pp. 5704–5707, Sep. 2004. doi: 10.1128/aem.70.9.5704-5707.2004 (cit. on p. 16).
- [69] H.-P. Fierobe, F. Mingardon, A. Mechaly, A. Bélaïch, M. T. Rincon, S. Pagès, R. Lamed, C. Tardif, J.-P. Bélaïch, and E. A. Bayer, "Action of designer cellulosomes on Homogeneous Versus Complex substrates", *Journal of Biological Chemistry*, vol. 280, no. 16, pp. 16 325–16 334, Feb. 2005. doi: 10.1074/jbc.m414449200 (cit. on p. 16).
- [70] S. Morais, Y. Barak, Y. Hadar, D. B. Wilson, Y. Shoham, R. Lamed, and E. A. Bayer, "Assembly of xylanases into designer cellulosomes promotes efficient hydrolysis of the xylan component of a natural recalcitrant cellulosic substrate", *mBio*, vol. 2, no. 6, e00233–11–e00233–11, 2011–11. doi: 10.1128/mbio.00233-11 (cit. on p. 16).
- [71] J. Stern, A. Kahn, Y. Vazana, M. Shamshoum, S. Morais, R. Lamed, and E. A. Bayer, "Significance of relative position of cellulases in designer cellulosomes for optimized cellulolysis", *PLOS ONE*, vol. 10, no. 5, E. A. Johnson, Ed., e0127326, May 2015. doi: 10.1371/journal.pone.0127326 (cit. on p. 16).
- [72] L. Davidi, S. Morais, L. Artzi, D. Knop, Y. Hadar, Y. Arfi, and E. A. Bayer, "Toward combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome", *Proceedings of the National Academy of Sciences*, vol. 113, no. 39, pp. 10 854–10 859, Sep. 2016. doi: 10.1073/pnas.1608012113 (cit. on p. 16).
- [73] J. Stern, L. Artzi, S. Morais, C. M. G. A. Fontes, and E. A. Bayer, "Carbohydrate depolymerization by intricate cellulosomal systems", in *Methods in Molecular Biology*, Springer New York, 2017, pp. 93–116. doi: 10.1007/978-1-4939-6899-2\_8 (cit. on p. 16).
-

- [74] C. You, S. Myung, and Y.-H. P. Zhang, "Facilitated substrate channeling in a self-assembled trifunctional enzyme complex", *Angewandte Chemie International Edition*, vol. 51, no. 35, pp. 8787–8790, Jul. 2012. doi: 10.1002/anie.201202441 (cit. on p. 16).
- [75] F. Liu, S. Banta, and W. Chen, "Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced NADH production", *Chemical Communications*, vol. 49, no. 36, p. 3766, 2013. doi: 10.1039/c3cc40454d (cit. on p. 16).
- [76] C. You and Y.-H. P. Zhang, "Annexation of a high-activity enzyme in a synthetic three-enzyme complex greatly decreases the degree of substrate channeling", *ACS Synthetic Biology*, vol. 3, no. 6, pp. 380–386, Dec. 2013. doi: 10.1021/sb4000993 (cit. on p. 16).
- [77] A. Szczupak, D. Aizik, S. Moraïs, Y. Vazana, Y. Barak, E. Bayer, and L. Alfonta, "The electrosome: A surface-displayed enzymatic cascade in a biofuel cell's anode and a high-density surface-displayed biocathodic enzyme", *Nanomaterials*, vol. 7, no. 7, p. 153, Jun. 2017. doi: 10.3390/nano7070153 (cit. on p. 16).
- [78] Z. Xiao, R. Storms, and A. Tsang, "Microplate-based filter paper assay to measure total cellulase activity", *Biotechnology and Bioengineering*, vol. 88, no. 7, pp. 832–837, 2004. doi: 10.1002/bit.20286 (cit. on p. 16).
- [79] C. Breuil and J. Saddler, "Comparison of the 3, 5-dinitrosalicylic acid and nelson-somogyi methods of assaying for reducing sugars and determining cellulase activity", *Enzyme and Microbial Technology*, vol. 7, no. 7, pp. 327–332, Jul. 1985. doi: 10.1016/0141-0229(85)90111-5 (cit. on p. 16, 17).
- [80] M. Dashtban, M. Maki, K. T. Leung, C. Mao, and W. Qin, "Cellulase activities in biomass conversion: Measurement methods and comparison", *Critical Reviews in Biotechnology*, vol. 30, no. 4, pp. 302–309, Sep. 2010. doi: 10.3109/07388551.2010.490938 (cit. on p. 17).
- [81] K.-J. Cheng, C. Stewart, D. Dinsdale, and J. Costerton, "Electron microscopy of bacteria involved in the digestion of plant cell walls", *Animal Feed Science and Technology*, vol. 10, no. 2-3, pp. 93–120, Jan. 1984. doi: 10.1016/0377-8401(84)90002-6 (cit. on p. 17).
- [82] P. Bubner, H. Plank, and B. Nidetzky, "Visualizing cellulase activity", *Biotechnology and Bioengineering*, vol. 110, no. 6, pp. 1529–1549, Mar. 2013. doi: 10.1002/bit.24884 (cit. on p. 17).
- [83] M. A. Nash, T. Verdorfer, and K. H. Malinowska, "Method of determining the degradation of cellulosic materials", WO2015091772 A1, Jun. 2015 (cit. on p. 17).
- [84] E. Florin, V. Moy, and H. Gaub, "Adhesion forces between individual ligand-receptor pairs", *Science*, vol. 264, no. 5157, pp. 415–417, Apr. 1994. doi: 10.1126/science.8153628 (cit. on p. 19).

- 
- [85] E. L. Florin, M. Rief, H. Lehmann, M. Ludwig, C. Dornmair, V. T. Moy, and H. E. Gaub, "Sensing specific molecular interactions with the atomic force microscope", *Biosensors and Bioelectronics*, vol. 10, no. 9-10, pp. 895–901, 1995 (cit. on p. 19).
- [86] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub, "Reversible Unfolding of Individual Titin Immunoglobulin Domains by AFM", *Science*, vol. 276, no. 5315, pp. 1109–1112, 1997. doi: 10.1126/science.276.5315.1109 (cit. on p. 19).
- [87] F. Oesterhelt, M. Rief, and H. E. Gaub, "Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water", *New J. Phys.*, vol. 1, no. 6, pp. 1–11, 1999. doi: 10.1088/1367-2630/1/1/006 (cit. on p. 19).
- [88] C. Yuan, A. Chen, P. Kolb, and V. T. Moy, "Energy landscape of streptavidin-biotin complexes measured by atomic force microscopy†", *Biochemistry*, vol. 39, no. 33, pp. 10 219–10 223, Aug. 2000. doi: 10.1021/bi992715o (cit. on p. 19).
- [89] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy", *Nat. Methods*, vol. 5, no. 6, pp. 491–505, 2008. doi: 10.1038/nmeth.1218 (cit. on p. 19).
- [90] M. Jobst, "Multiplexed single molecule observation and manipulation of engineered biomolecules", 2018 (cit. on p. 19).
- [91] W. Ott, M. A. Jobst, C. Schoeler, H. E. Gaub, and M. A. Nash, "Single-molecule force spectroscopy on polyproteins and receptor–ligand complexes: The current toolbox", *J. Struct. Biol.*, vol. 197, no. 1, pp. 3–12, 2017. doi: 10.1016/j.jsb.2016.02.011 (cit. on p. 19).
- [92] G. Binnig, C. F. Quate, and C. Gerber, "Atomic force microscope", *Physical Review Letters*, vol. 56, no. 9, pp. 930–933, Mar. 1986. doi: 10.1103/physrevlett.56.930 (cit. on p. 19).
- [93] S. Suresh, "Elastic clues in cancer detection", *Nature Nanotechnology*, vol. 2, no. 12, pp. 748–749, Dec. 2007. doi: 10.1038/nnano.2007.397 (cit. on p. 19).
- [94] C. Roduit, S. Sekatski, G. Dietler, S. Catsicas, F. Lafont, and S. Kasas, "Stiffness tomography by atomic force microscopy", *Biophysical Journal*, vol. 97, no. 2, pp. 674–677, Jul. 2009. doi: 10.1016/j.bpj.2009.05.010 (cit. on p. 19).
- [95] F. M. Hecht, J. Rheinlaender, N. Schierbaum, W. H. Goldmann, B. Fabry, and T. E. Schäffer, "Imaging viscoelastic properties of live cells by AFM: Power-law rheology on the nanoscale", *Soft Matter*, vol. 11, no. 23, pp. 4584–4591, 2015. doi: 10.1039/c4sm02718c (cit. on p. 19).
- [96] V. Moy, E. Florin, and H. Gaub, "Intermolecular forces and energies between ligands and receptors", *Science*, vol. 266, no. 5183, pp. 257–259, Oct. 1994. doi: 10.1126/science.7939660 (cit. on p. 19).
- [97] M. Grandbois, "How strong is a covalent bond?", *Science*, vol. 283, no. 5408, pp. 1727–1730, Mar. 1999. doi: 10.1126/science.283.5408.1727 (cit. on p. 19).
-

- [98] L. Schmitt, M. Ludwig, H. E. Gaub, and R. Tampé, "A metal-chelating microscopy tip as a new toolbox for single-molecule experiments by atomic force microscopy", *Biophysical Journal*, vol. 78, no. 6, pp. 3275–3285, Jun. 2000. doi: 10.1016/S0006-3495(00)76863-9 (cit. on p. 19).
- [99] O. Kratky and G. Porod, "Röntgenuntersuchung gelöster fadenmoleküle", *Recl. Trav. Chim. Pays-Bas*, vol. 68, no. 12, pp. 1106–1122, 1949. doi: 10.1002/recl.19490681203 (cit. on p. 22).
- [100] S. Brinkers, H. R. C. Dietrich, F. H. de Groote, I. T. Young, and B. Rieger, "The persistence length of double stranded DNA determined using dark field tethered particle motion", *The Journal of Chemical Physics*, vol. 130, no. 21, p. 215 105, Jun. 2009. doi: 10.1063/1.3142699 (cit. on p. 22).
- [101] H. Dietz and M. Rief, "Exploring the energy landscape of GFP by single-molecule mechanical experiments", *Proceedings of the National Academy of Sciences*, vol. 101, no. 46, pp. 16 192–16 197, Nov. 2004. doi: 10.1073/pnas.0404549101 (cit. on p. 22).
- [102] J. F. Marko and E. D. Siggia, "Stretching DNA", *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, Dec. 1995. doi: 10.1021/ma00130a008 (cit. on p. 22).
- [103] R. Petrosyan, "Improved approximations for some polymer extension models", *Rheologica Acta*, vol. 56, no. 1, pp. 21–26, Jan. 2017. doi: 10.1007/s00397-016-0977-9 (cit. on p. 22).
- [104] T. Hugel, M. Rief, M. Seitz, H. E. Gaub, and R. R. Netz, "Highly Stretched Single Polymers: Atomic-Force-Microscope Experiments Versus Ab-initio Theory", *Phys. Rev. Lett.*, vol. 94, no. 4, pp. 048301-1–048301-4, 2005. doi: 10.1103/PhysRevLett.94.048301 (cit. on p. 22).
- [105] G. I. Bell, "Models for the specific adhesion of cells to cells", *Science*, vol. 200, no. 4342, pp. 618–627, 1978. doi: 10.1126/science.347575 (cit. on p. 23).
- [106] E. Evans and K. Ritchie, "Dynamic strength of molecular adhesion bonds", *Biophys. J.*, vol. 72, no. 4, pp. 1541–1555, 1997. doi: 10.1016/S0006-3495(97)78802-7 (cit. on p. 23).
- [107] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, "Molecular dynamics study of unbinding of the avidin-biotin complex", *Biophys. J.*, vol. 72, no. 4, pp. 1568–1581, 1997. doi: 10.1016/S0006-3495(97)78804-0 (cit. on p. 23).
- [108] U. Brand, S. Gao, W. Engl, T. Sulzbach, S. W. Stahl, L. F. Milles, V. Nesterov, and Z. Li, "Comparing AFM cantilever stiffness measured using the thermal vibration and the improved thermal vibration methods with that of an SI traceable method based on MEMS", *Measurement Science and Technology*, vol. 28, no. 3, p. 034 010, Jan. 2017. doi: 10.1088/1361-6501/28/3/034010 (cit. on p. 23).
- [109] J. L. Hutter and J. Bechhoefer, "Calibration of atomic-force microscope tips", *Review of Scientific Instruments*, vol. 64, no. 7, pp. 1868–1873, Jul. 1993. doi: 10.1063/1.1143970 (cit. on p. 23).

- 
- [110] H. .-.-J. Butt and M. Jaschke, "Calculation of thermal noise in atomic force microscopy", *Nanotechnology*, vol. 6, no. 1, pp. 1–7, Jan. 1995. doi: 10.1088/0957-4484/6/1/001 (cit. on p. 24).
- [111] C. A. Clifford and M. P. Seah, "The determination of atomic force microscope cantilever spring constants via dimensional methods for nanomechanical analysis", *Nanotechnology*, vol. 16, no. 9, pp. 1666–1680, Jul. 2005. doi: 10.1088/0957-4484/16/9/044 (cit. on p. 24).
- [112] R. Proksch, T. E. Schäffer, J. P. Cleveland, R. C. Callahan, and M. B. Viani, "Finite optical spot size and position corrections in thermal spring constant calibration", *Nanotechnology*, vol. 15, no. 9, pp. 1344–1350, Aug. 2004. doi: 10.1088/0957-4484/15/9/039 (cit. on p. 24).
- [113] T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash, and H. E. Gaub, "Combining in vitro and in silico Single-Molecule force spectroscopy to characterize and tune cellulosomal scaffoldin mechanics", *J. Am. Chem. Soc.*, 2017. doi: 10.1021/jacs.7b07574 (cit. on p. 30).
- [114] T. Verdorfer and H. E. Gaub, "Ligand binding stabilizes cellulosomal cohesins as revealed by AFM-based single-molecule force spectroscopy", *Scientific Reports*, vol. 8, no. 1, Jul. 2018. doi: 10.1038/s41598-018-27085-x (cit. on p. 70).
- [115] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, and H. E. Gaub, "From genes to protein mechanics on a chip", *Nat. Methods*, vol. 11, no. 11, pp. 1127–1130, 2014. doi: 10.1038/nmeth.3099 (cit. on p. 86).
- [116] C. Schoeler, T. Verdorfer, H. E. Gaub, and M. A. Nash, "Biasing effects of receptor-ligand complexes on protein-unfolding statistics", *Physical Review E*, vol. 94, no. 4, Oct. 2016. doi: 10.1103/physreve.94.042412 (cit. on p. 108).
- [117] K. H. Malinowska, T. Verdorfer, A. Meinhold, L. F. Milles, V. Funk, H. E. Gaub, and M. A. Nash, "Redox-initiated hydrogel system for detection and real-time imaging of cellulolytic enzyme activity", *ChemSusChem*, vol. 7, no. 10, pp. 2825–2831, Aug. 2014. doi: 10.1002/cssc.201402428 (cit. on p. 118).
- [118] K. H. Malinowska, T. Rind, T. Verdorfer, H. E. Gaub, and M. A. Nash, "Quantifying synergy, thermostability, and targeting of cellulolytic enzymes and cellulosomes with polymerization-based amplification", *Analytical Chemistry*, vol. 87, no. 14, pp. 7133–7140, Jun. 2015. doi: 10.1021/acs.analchem.5b00936 (cit. on p. 132).
-



# Acknowledgements

I would like to thank everyone who has helped and supported me in the course of my doctoral studies, specifically a number of people without whom my work and this thesis would not have been possible:

First and foremost, **Prof. Dr. Hermann Gaub** for providing an awesome research environment and great scientific guidance over the years. I am very grateful for all the (bio-)physics knowledge and experience I gained during my time under your wings.

**Prof. Dr. Michael Nash** for taking me into the Nashcats group back in 2012, for being my mentor and motivator for many years, and for conceiving many of the projects that now comprise this thesis.

Very special thanks go my (office) companions **Ellis Durner, Lukas Milles, Markus Jobst, Wolfgang Ott** and also **Thomas Nicolaus**. Thank you very much for all the shared fun, pain, friendship, work, science, (sometimes heated) discussions, soft- and hardware developments, publications, proof-readings, after-work beers and of course the awesome trips to Südtirol. I am certain our friendship will last past our crazy office and I am looking forward to the trips to come. Also, thank you, Markus, for providing this beautiful L<sup>A</sup>T<sub>E</sub>X template, which is originally based on a template by Jordan Suchow.

**Constantin Schöler** for being one of my closest friends from day one and his support even through hard times. Furthermore, for teaching me that phoning people isn't as bad as I thought it is and for being the one guy in the office I could have discussions about cars with.

I also want to thank the rest of the **Gambicrew** for all the helpful and fun scientific discussions, all the support and transferred knowledge, and of course the great and unique working atmosphere.

Special thanks go to **Rafael Bernardi** from the UIUC for his enthusiasm and great work on *in silico* force spectroscopy. Some of the awesome results that led to this thesis would not have been possible without him.

Furthermore I would like to thank **Frank "Knarf" Kofler** and **Andrey Kuzmin** for all the support and fun over the last few years. Our weekly gaming tradition has surely helped keeping me sane during these last few years. I hope we can keep it alive for as long as possible, bois!

I would like to express my deep and sincere gratitude to my many friends and of course my family for their continuous and unparalleled love, help and support. I would not be where I am today without you.

Finally, I want to thank **Lisa Horle** for her support and encouragement during these last years. Having you on my side is worth more than I can express on paper.